

SONDERFORSCHUNGSBEREICH 504

Rationalitätskonzepte,
Entscheidungsverhalten und
ökonomische Modellierung

No. 04-16

**Die angemessene Erfassung der Stärke und
Richtung von Anreizen durch soziale Erwünschtheit**

Volker Stocké*
and Christian Hunkler**

May 2004

Financial support from the Deutsche Forschungsgemeinschaft, SFB 504, at the University of Mannheim, is gratefully acknowledged.

*Sonderforschungsbereich 504, email: vstocke@rumms.uni-mannheim.de

**Sonderforschungsbereich 504, email:



Universität Mannheim
L 13,15
68131 Mannheim

Abstract¹

The study investigates the social desirability of four characteristics often asked as self-descriptive questions in surveys. These are the intensity of watching television, the consumption of alcohol, the frequency of past blood donations and the degree of happiness in life. We compare the validity of three differently elaborated operationalizations of the magnitude and direction of the perceived incentives for social desirable response behavior. The standard method only accounts for the social desirability of a strong occurrence of the respective attribute, however, the second procedure takes differences between strong and weak occurrences into account, and the third implementation considers additionally the evaluation of a medium occurrence. Using data from a local survey based on a random probability sample, first, we find a substantial proportion of the respondents reporting u-shaped desirability profiles for the consumption of alcohol and intensity of watching television. For these topics only the third and most sophisticated operationalization is an appropriate measure of the potential distortions caused by social desirability. Secondly, monotonic desirability profiles are observed for past blood donations and happiness in life. Here, the additional effort necessary for the global difference scores as the second most elaborated measure, improves considerably the magnitude of incentives from social desirability which can be identified. Contemporary research has almost exclusively applied the standard method; however, our investigation suggests that this is under no conditions an adequate measure of the strength and the direction of potential social desirability bias.

1 Einleitung

Antwortverzerrungen durch soziale Erwünschtheit sind eine erhebliche Bedrohung der Gültigkeit von Umfragedaten. Belege für die Wirksamkeit dieser Art von Response-Effects wurden beispielsweise in Umfragen über die Kirchgangshäufigkeit und die psychische Gesundheit der Befragten sowie über deren Drogen- und Alkoholkonsum gefunden (*Aquilino* 1998; *Epstein* et al. 2001; *Jobe* et al. 1997; *Presser* und *Stinson* 1998). Da es sich beim „social desirability bias“ (SD-Bias) um systematische Antwortverzerrungen in eine bestimmte Richtung handelt, muss als Ergebnis mit verzerrten Randverteilungen der abgefragten Einstellungen und faktischen Merkmale der Befragten gerechnet werden. Diese Antwortverzerrungen können dann durch unterschiedliche statistische Verfahren nachträglich korrigiert werden, wenn ihre Stärke und inhaltliche Ausrichtung möglichst genau und verlässlich vorhergesagt werden können.

¹ Wir danken ganz herzlich *Hartmut Esser* und den Gutachtern dieser Zeitschrift für hilfreiche Kommentare und Anregungen. *Diana Schirowski* und *Jason Rosow* waren eine große Unterstützung bei der Erstellung des Manuskriptes. Der Beitrag wurde durch finanzielle Förderung der Deutschen Forschungsgemeinschaft an den Sonderforschungsbereich 504 der Universität Mannheim ermöglicht.

Im Rahmen der Rational-Choice Theorie (RCT) des Befragtenverhaltens, aber auch in anderen Ansätzen, die den SD-Bias als das Ergebnis einer an die soziale Umwelt gerichteten „Impression Management“-Strategie erklären, werden die Erwünschtheitswahrnehmungen der Befragten als eine zentrale Entstehungsbedingung des SD-Bias angesehen (*Esser* 1991, *Stocké* 2004). Demnach verwenden die Befragten ihre Wahrnehmungen über die Erwünschtheit oder „Trait Desirability“ (TD) eines Merkmals als Information darüber, welche Wertungen und Sanktionen durch die soziale Umwelt im Interview bei einer Selbstbeschreibung dieses Merkmals durch ihr Antwortverhalten durchschnittlich erwartet werden müssen. Die Erwünschtheitswahrnehmungen im konkreten Befragungskontext ergeben sich einerseits durch bewertende Reaktionen der Interviewer im vorangegangenen Teil der Befragung und durch auf der Grundlage von Interviewermerkmalen antizipierten Erwartungshaltungen. Andererseits verwenden die Befragten ihr Wissen über gesellschaftliche Normen, um Erwartungen über die Konsequenzen ihres Antwortverhaltens zu generieren. Aus der Perspektive der RCT streben Befragte durch ihre Antworten Verhaltensbestätigung, die Vermeidung negativer Sanktionen und insgesamt soziale Anerkennung an. Die Erwünschtheitswahrnehmungen sind eine zentrale *kognitive* Grundlage für die von den Befragten angenommene Eignung verschiedener Antwortalternativen zur Realisierung dieser Ziele. Eine valide Erfassung dieser Wahrnehmungen ist damit eine notwendige, aber nicht hinreichende Vorbedingung für eine genaue Prognose des SD-Bias.

Aus der theoretischen Logik der RCT prognostizieren die Erwünschtheitswahrnehmungen die Stärke und Richtung von Antwortverzerrungen durch soziale Erwünschtheit nur dann, wenn die Antworten durch anwesende Personen wahrgenommen werden können und außerdem ein hinreichend starkes Bedürfnis nach sozialer Anerkennung vorliegt (*Stocké* 2004). Es wird also von einer nichtkompensatorischen, multiplikativen Verknüpfung dieser drei Entstehungsbedingungen ausgegangen. Liegen keine externen Validierungsdaten vor, so kann die Erklärungskraft dieser Entstehungsbedingungen des SD-Bias für das Antwortverhalten und insbesondere das prognostizierte interaktive Zusammenspiel dieser Bedingungen als Beleg für die Existenz von Antwortverzerrungen herangezogen werden.

Eine besonders schwerwiegende Beeinträchtigung der Verlässlichkeit von Umfragedaten liegt dann vor, wenn sich der SD-Bias verschieden stark oder gar in unterschiedlicher Richtung auf das Antwortverhalten durch bestimmte Merkmale definierter Teilgruppen der Befragten auswirkt. Unter diesen Umständen muss damit gerechnet werden, dass beobachtete Zusammenhänge zwischen dem betreffenden Erhebungsmerkmal und der entsprechenden Gruppenzugehörigkeit reine Artefakte der (unterschiedlichen) Wirksamkeit sozialer Erwünschtheit sind. Eine mögliche Ursache für eine gruppenspezifische Differenzierung in der Stärke und Richtung des SD-Bias ist, dass sich die Erwünschtheitswahrnehmungen entlang der Gruppengrenzen unterscheidet. Der Grund für solche Unterschiede kann in der Einbettung in unterschiedliche normative Bezugsgruppen und deren Prägekraft für die als gültig angesehenen Bewertungskriterien gesehen werden (*Singer* 1981). So kann etwa das

empirische Ergebnis, wonach Männer in Umfragen bei der Angabe der Anzahl ihrer Geschlechtspartner übertreiben, während Frauen in dieser Hinsicht zu geringe Angaben machen, als Hinweis auf die Existenz geschlechtsspezifischer Normen über das angemessene Ausmaß sexueller Promiskuität gewertet werden (*Smith* 1992). Ob diese und andere Unterschiede in der Richtung des SD-Bias tatsächlich auf Gruppenunterschiede in der Erwünschtheitswahrnehmung zurückgehen, wurde jedoch bisher kaum untersucht. Eine solche Analyse erfordert ein gültiges und zuverlässiges Instrument zur Erfassung der von den Befragten subjektiv wahrgenommenen Erwünschtheit des betreffenden Merkmals.

Bei der bisherigen Erforschung der sozialen Erwünschtheit unterschiedlicher, in Umfragen durch Selbstbeschreibungsfragen erfasster Merkmale wird in der Regel die antizipierte Bewertung der Anwesenheit des betreffenden Merkmals erfasst. So haben beispielsweise die Befragten in der Untersuchung von *Wyner* (1980) angegeben, für wie sozial erwünscht oder unerwünscht sie verschiedene Straftaten, wie etwa Raub oder Einbruch, ansehen. Diese Vorgehensweise bezieht sich bei Merkmalen mit einer kontinuierlich variierenden Ausprägungsstärke – wie etwa dem Ausmaß des Alkoholkonsums – auf die wahrgenommene gesellschaftliche Bewertung einer starken Merkmalsausprägung. Diese Standardoperationalisierung des TD-Konstruktes setzt implizit voraus, dass eine schwache bzw. fehlende Ausprägung des untersuchten Merkmals von den Informanten als neutral bewertet angesehen wird (vgl. hierzu noch Abschnitt 2.1). Diese Annahme wurde in der bisherigen Forschung aber kaum empirisch geprüft. Zu diesem Zweck ist es notwendig, auch die wahrgenommene Erwünschtheit der Abwesenheit bzw. einer schwachen Ausprägung des betreffenden Merkmals zu erfassen. Bei dieser differenzierteren Operationalisierung des TD-Konstruktes wird bei Merkmalen mit kontinuierlich variierender Ausprägungsstärke eine weitere und unseren Wissens nach bisher völlig ungetestete Annahme gemacht. Diese Annahme besteht darin, dass die Erwünschtheitswahrnehmungen der Befragten mit zunehmender Intensität der Merkmalsausprägung einen streng monoton ansteigenden oder abnehmenden Verlauf nimmt. Im Widerspruch hierzu sind jedoch soziale Normen denkbar, wonach einer mittleren Merkmalsstärke oder gar den beiden extremen Merkmalsausprägungen das höchste Ausmaß an sozialer Erwünschtheit zukommt. Unter diesen Umständen ist weder die Standardoperationalisierung noch die differenzierte Erfassung der TD mit zwei Messpunkten eine angemessene Repräsentation der Stärke und Richtung der Anreize durch soziale Erwünschtheit. Ob dies der Fall ist, kann nur durch eine noch weiter differenzierte Erfassung der Erwünschtheitswahrnehmungen für eine schwache, mittlere und starke Merkmalsausprägung überprüft werden.

Der vorliegende Beitrag analysiert bei vier häufig in Umfragen erfassten Merkmalen, wie differenziert bei einer angemessenen Erfassung der jeweils vorliegenden Stärke und Richtung der Anreize zu sozial erwünschtem Antwortverhalten vorgegangen werden muss. Die untersuchten Merkmale sind das Ausmaß des Lebensglücks und die Häufigkeit von Blutspenden der Befragten, sowie die Intensität ihres Alkohol- und Fernsehkonsums. Es wird

analysiert, ob und bei welchen dieser Befragungsthemen den drei möglichen Operationalisierungen des TD-Konzeptes konvergente Validität zukommt.

2 Forschungsstand

Im folgenden Abschnitt werden zuerst die unterschiedlichen Versionen des TD-Konzeptes sowie verschiedene Möglichkeiten ihrer Operationalisierung genauer diskutiert. Im zweiten Abschnitt wird dann die verfügbare Evidenz für die Prognosekraft dieser Indikatoren der wahrgenommenen sozialen Erwünschtheit für die Entstehung des SD-Bias dargestellt. Der dritte Abschnitt beschreibt den Kenntnisstand über die Stärke und Richtung des zu erwartenden SD-Bias bei den im vorliegenden Beitrag untersuchten Befragungsthemen.

2.1 Operationalisierungen der Trait-Desirability

Die soziale Erwünschtheit eines Merkmals und die damit verbundene Sanktionserwartung der Befragten kann entweder in der Form „persönlicher“ oder „extrinsisch evaluierter“ Bewertungen operationalisiert werden. Im ersten Fall geben die Informanten an, für wie erwünscht sie persönlich bestimmte Eigenschaften oder Verhaltensweisen einer Person ansehen (vgl. beispielsweise: *Meleddu* und *Guicciardi* 1998; *Smith* und *Smith* 1993). Die aggregierten Angaben der Informanten über ihre eigenen Erwünschtheitsurteile können dann als Indikator für das normative Klima in der befragten Stichprobe angesehen werden. Dieser „objektive“ Indikator für die von einem Befragten bei der Selbstzuschreibung des betreffenden Merkmals zu erwartenden positiven oder negativen Sanktionen ist dann valide, wenn die so erfassten Urteile mit den von den Befragten subjektiv erwarteten Bewertungsreaktionen übereinstimmt. Dies ist empirisch nachweisbar nicht immer der Fall (*Crott* und *Roßbrucker* 1974). Die zweite mögliche Art der Operationalisierung des TD-Konstruktes, die „extrinsisch“ evaluierte oder fremdeingeschätzte soziale Erwünschtheit, zielt direkt auf die von Befragten subjektiv wahrgenommenen, in der Gesellschaft dominanten Bewertungsstandards (vgl. beispielsweise: *Crott* und *Roßbrucker* 1974; *Edwards* 1957; *Nicotera* 1996). Die Befragten werden hierbei um ihre Einschätzung darüber gebeten, welche Bewertungen oder Reaktionen ein bestimmtes Merkmal typischerweise hervorruft.

Bei der empirischen Erfassung beider Varianten der Operationalisierung des TD-Konstruktes sind weiterhin unterschiedlich differenzierte und damit verschieden aufwendige Vorgehensweisen denkbar. Bei der *ersten* und in der derzeitigen Forschung fast ausschließlich angewandten Standardmethode wird die von den Befragten wahrgenommene Erwünschtheit des Vorliegens oder, im Falle kontinuierlich variierender Eigenschaften, einer starken Ausprägung des untersuchten Merkmals erfasst. Es werden typischerweise bipolare Antwortskalen herangezogen, deren Mittelpunkt eine neutrale Erwünschtheitswahrnehmung repräsentiert. Das Ausmaß und die Richtung der Abweichung der antizipierten Merkmals-

bewertung vom neutralen Mittelpunkt der Skala wird als Indikator für die bestehenden Anreize zur Selbstzuschreibung bzw. Leugnung einer starken Merkmalsausprägung angesehen. Aus der Perspektive der RCT ist diese Vorgehensweise nur dann angemessen, wenn die schwache Merkmalsausprägung von den Befragten als neutral bewertet angesehen wird. Dies lässt sich aus der Annahme begründen, dass die Befragten die Differenz in der Bewertung der extremen Pole des Merkmalskontinuums als Grundlage für ihre Wahrnehmungen der Anreize durch soziale Erwünschtheit heranziehen (*Stocké* 2004). Einer positiven oder negativen Erwünschtheitseinschätzung einer starken Merkmalsausprägung alleine kommt demnach nur dann Konstruktvalidität zu, wenn diese Einschätzung mit der Differenz der bewerteten Erwünschtheit einer starken und schwachen Merkmalsausprägung übereinstimmt. Ist dies wegen der Verletzung der Neutralitätsannahme nicht erfüllt, so repräsentiert die Standardoperationalisierung eine Über- oder Unterschätzung der Stärke und im schwerwiegendsten Fall eine Fehleinschätzung der Richtung der tatsächlich vorliegenden Anreize durch soziale Erwünschtheit. Eine Unterschätzung liegt dann vor, wenn die Erwünschtheitswahrnehmung einer starken und schwachen Merkmalsausprägung in unterschiedlicher Richtung von einer neutralen Bewertung abweicht: die Erwünschtheitsunterschiede zwischen den beiden Antwortmöglichkeiten sind dann in Realität größer als die bei einer starken Merkmalsausprägung erfassten Abweichungen vom neutralen Mittelpunkt der Skala. Weicht dagegen die soziale Erwünschtheit der schwachen Merkmalsausprägung in die gleiche Richtung wie die einer starken Ausprägung von einer neutralen Bewertung ab, so führt die Standardmethode entweder zu einer Überschätzung des Verzerrungspotentials durch soziale Erwünschtheit oder zu falschen Schlussfolgerungen über die Richtung des Verzerrungspotentials durch den SD-Bias. Eine Überschätzung liegt dann vor, wenn eine schwache Merkmalsausprägung weniger erwünscht bzw. unerwünscht als eine starke Ausprägung eingeschätzt wird: die realen Erwünschtheitsunterschiede zwischen der starken und schwachen Merkmalsausprägung sind dann geringer als die bei einer starken Ausprägung beobachtete Abweichung vom neutralen Skalenmittelpunkt. Wird die schwache Merkmalsausprägung dagegen sogar als erwünschter bzw. unerwünschter angesehen als dies bei der erfassten Bewertung einer starken Ausprägung der Fall ist, so nehmen die Befragten den Implikationen der Standardoperationalisierung entgegengesetzte Anreize zu sozial erwünschtem Antwortverhalten wahr.

Bei der *zweiten*, in der empirischen Forschung bisher kaum angewandten Vorgehensweise wird die Annahme aufgegeben, dass eine schwache Merkmalsausprägung notwendigerweise neutral bewertet wird. Bei dieser differenzierteren Operationalisierung des TD-Konstruktes wird neben der Bewertung einer starken auch die einer schwachen Merkmalsausprägung explizit erfasst und die Differenz zwischen beiden Erwünschtheitswahrnehmungen als Indikator für die Stärke und Richtung der Anreize zum SD-Bias herangezogen (als Beispiel vgl.: *Stocké* 2004). Bei dieser Operationalisierung mit Differenzwerten wird, ebenso wie bei der Standardoperationalisierung des TD-Konstruktes, vorausgesetzt, dass die Erwünschtheits-

wahrnehmungen mit steigender Ausprägungsstärke der Merkmale monoton zunehmen oder abnehmen. Ist diese Annahme nicht erfüllt, so führen die damit verbundenen u-förmigen oder umgekehrt u-förmigen Erwünschtheitsprofile nicht zu Antwortverzerrungen in eine bestimmte Richtung, sondern zu einer Konzentrierung der Angaben entweder am Mittelpunkt oder an beiden Extrempolen des Merkmalskontinuums. Derartige nicht-monotone Verläufe der sozialen Erwünschtheit über das Merkmalskontinuum lassen sich nur durch die zusätzliche Erfassung der antizipierten Bewertung einer mittelstarken Merkmalsausprägung identifizieren. Da uns keine empirische Anwendung dieser *dritten*, bereichsspezifischen und daher differenziertesten Operationalisierung des TD-Konstruktes bekannt ist, ist derzeit unklar, ob der hierbei zusätzlich notwendige empirische Aufwand gerechtfertigt ist.

Insgesamt kann mit einer vollständigen konvergenten Validität der drei Operationalisierungen nur dann gerechnet werden, wenn sowohl die Neutralitätsannahme als auch die Monotonieannahme bei allen Befragungsthemen empirisch erfüllt sind. Ist die Neutralitätsannahme verletzt, so kommt der Differenz- sowie der bereichsspezifischen Operationalisierung weiterhin konvergente Validität zu, die Standardoperationalisierung führt dann allerdings zu einer fehlerhaften Prognose der vorliegenden Stärke und eventuell auch der Richtung von Anreizen zu sozial erwünschtem Antwortverhalten. Liegen dagegen bei einem relevanten Teil der Befragten nicht monotone Erwünschtheitsverläufe vor, so sind weder die Standardoperationalisierung noch die Verwendung von globalen Differenzwerten eine angemessene Erfassung der von den Befragten wahrgenommenen Anreize durch soziale Erwünschtheit. Bei gleichzeitiger Verletzung der Neutralitäts- als auch die Monotonieannahme liegt somit das geringste Ausmaß an konvergenter Validität der drei Messkonzepte vor. Für die vier in dieser Arbeit analysierten Befragungsthemen wurde unseres Wissens nach bisher keine der beiden Annahmen getestet.

Als aggregierte Maßzahl für den bei einem bestimmten Befragungsthema drohenden SD-Bias kann bei der Standardoperationalisierung des TD-Konstruktes und bei der Verwendung globaler Differenzwerte der Durchschnitt der individuellen Skalenwerte der Informanten herangezogen werden. Bei der Verwendung bereichsspezifischer Differenzwerte ist diese Maßzahl durch die Summe der aggregierten Erwünschtheitsunterschiede zwischen einer starken und mittleren sowie mittleren und schwachen Merkmalsausprägung definiert. Bei allen drei Operationalisierungen gibt der Betrag des resultierenden Indikators die Stärke und das Vorzeichen die Richtung des auf der Aggregatebene erwartbaren SD-Bias an. Eine auf der Individualebene der Befragten möglicherweise vorliegende Heterogenität hinsichtlich der Richtung wahrgenommener Anreize durch soziale Erwünschtheit bleibt hierbei jedoch unberücksichtigt. Da sich solche Unterschiede in der Richtung der Anreizwahrnehmungen auf der Aggregatebene gegenseitig aufheben, kommt es unter diesen Umständen zu einer Unterschätzung des Einflusspotentials durch SD-Bias. Obwohl eine solche Unterschätzung für die beobachtete Randverteilung der untersuchten Merkmale ohne Konsequenzen ist, können Unterschiede in der Richtung von Antwortverzerrungen auf der Ebene von Sub-

gruppen der Befragten zu Abweichungen der beobachteten von den Zusammenhängen zwischen dem untersuchten Merkmal und den relevanten Gruppierungsmerkmalen bewirken. Ob eine solche Heterogenität hinsichtlich der Richtung der Anreizwahrnehmungen vorliegt, kann bei jeder der drei Operationalisierungen des SD-Konstruktes durch eine nach dem Vorzeichen der individuellen Anreizwahrnehmungen differenzierten Analyse kontrolliert werden.

2.2 Bedeutung der Erwünschtheitswahrnehmungen für die Erklärung des SD-Bias

In einer Reihe von Studien wurde die soziale Erwünschtheit unterschiedlicher Merkmale mit der Absicht einer Vorhersage von Antwortverzerrungen durch soziale Erwünschtheit erfasst. So haben vorliegende Studien mit britischen und australischen Studenten gezeigt, dass apathische oder arrogante Personen als negativ, Loyalität und Vorurteilslosigkeit dagegen als sehr positiv angesehen werden (*Bochner* und *van Zyl* 1984; *Hampson* et al. 1987). In beiden Studien wurde nur die Bewertung des Vorliegens der jeweiligen Eigenschaft auf bipolaren Skalen abgefragt. In der Untersuchung von *Hampson* und Kollegen wurde erfasst, wie erwünscht die betreffende Eigenschaft für eine Person im allgemeinen ist, während in der Studie von *Bochner* und *van Zyl* gefragt wurde, für wie wünschbar die Informanten die analysierte Eigenschaft bei Familienmitgliedern, Freunden und Bekannten halten. Bei beiden Studien wird also die extrinsisch evaluierte Erwünschtheit mithilfe der Standardoperationalisierung eines Messpunktes erfasst.

Weitere Forschungsarbeiten zeigen, dass sich die soziale Erwünschtheit bei einigen Befragungsthemen deutlich zwischen unterschiedlichen Befragtengruppen unterscheidet. *Mecklenbräuker* und Kollegen (1994) ließen die Eigenschaften „listig“ und „willig“ auf einer bipolaren Skala mit neutralem Mittelpunkt von einer Stichprobe deutscher Studenten bewerten. Dabei unterschieden sie zwischen der Einschätzung der Wünschbarkeit bei einem gleichaltrigen Mann bzw. bei einer gleichaltrigen Frau. Zusätzlich wurden die Bewertungen nach dem Geschlecht der Informanten differenziert berichtet. Es zeigte sich, dass die Eigenschaft „listig“ bei Personen jeglichen Geschlechts von Männern als erwünschter eingeschätzt wird als von Frauen. Frauen bewerten das Merkmal „willig“ bei einer anderen Frau als negativ, bei einem gleichaltrigen Mann aber tendenziell als positiv. Auch in dieser Untersuchung wurde jeweils nur eine starke Ausprägung der jeweiligen Eigenschaft bewertet.

Dass sich die Erwünschtheitswahrnehmungen auch auf das Antwortverhalten der Befragten auswirken, wurde für eine Reihe von Befragungsthemen überprüft. Es konnte beispielsweise gezeigt werden, dass sich Befragten zunehmend durch ihr Antwortverhalten als argumentative Persönlichkeit charakterisieren, wenn sie diese Eigenschaft gesellschaftlich immer mehr als erwünscht einschätzen (*Nicotera* 1996). Die Teilnehmer der Studie haben die Items der „Infante and Rancer Argumentativeness Scale“ auf einer fünf-stufigen Antwortskala bewer-

tet, deren Endpunkte mit „nicht gesellschaftlich akzeptiert“ und „gesellschaftlich akzeptiert“ verankert waren. Die resultierenden fremdeingeschätzten Erwünschtheitswahrnehmungen haben sich als signifikant mit dem Antwortverhalten der Befragten auf der „Argumentativeness Scale“ korreliert erweisen ($r = .26$).

Ähnliche Zusammenhänge zwischen den Erwünschtheitswahrnehmungen und dem Antwortverhalten finden sich in Befragungen über das Vorliegen von Symptomen psychischer Störungen und bei der Erfassung einer großen Anzahl unterschiedlicher Persönlichkeitseigenschaften. In allen diesen Studien wird ebenfalls die Standardoperationalisierung des TD-Konstruktes herangezogen und somit nur das Vorliegen des jeweils analysierten Symptoms bewertet. *Phillips* und *Clancy* (1970) sowie *Edwards* (1957) lassen die Wünschbarkeit der Symptome auf einer unipolaren neun-stufigen Skala einschätzen: je höher die Erwünschtheit, desto höher der Skalenwert. *Phillips* und *Clancy* (1970) haben einen signifikant positiven Zusammenhang zwischen der Erwünschtheitseinstufung von Symptomen psychischer Gesundheit und der Anzahl berichteter Symptome des „22-Item Mental Health“ Indexes festgestellt: je höher die geschätzte Erwünschtheit, desto mehr Symptome werden berichtet. *Gove* und *Geerken* (1977) verwenden eine bipolare, durchgängig verbal verankerte Antwortskala, die zwischen extrem, ziemlich und etwas (un-) erwünscht differenziert. Auch hier wurde ein positiver Zusammenhang zwischen den Erwünschtheitsurteilen und der Anzahl der berichteten psychiatrischen Symptome gefunden. *Huang* und Kollegen (1998) finden vergleichbare Ergebnisse über den Zusammenhang zwischen den Erwünschtheitswahrnehmungen der Befragten und deren Antwortverhalten. In einer weiteren Studie kann gezeigt werden, dass nicht nur ein Zusammenhang zwischen den Erwünschtheitswahrnehmungen und dem Antwortverhalten besteht, sondern dass diese tatsächlich Antwortverzerrungen vorhersagen. So wurde durch Abgleich der Befragtenangaben über die Anzahl ihrer bisherigen Verhaftungen mit Informationen aus Polizeiakten gezeigt, dass die Umfrageteilnehmer bei der Angabe der Häufigkeit dieser Ereignisse in der Vergangenheit teilweise über- und teilweise untertrieben haben (*Wyner* 1980). Die Stärke und vor allem die Richtung dieser Antwortverzerrungen konnte durch die Unterschiede der von den Befragten wahrgenommenen gesellschaftlichen Bewertung des untersuchten Merkmals vorhergesagt werden.

Im Rahmen der Theorie rationalen Befragtenverhaltens wird vorhergesagt, dass sich die Erwünschtheitswahrnehmungen in Interaktion mit dem Bedürfnis nach sozialer Anerkennung und der Privatheit der Antwortsituation auf die Angaben der Befragten auswirken. Diese Annahme wurde in einer Studie über die Affektivität der Befragten überprüft (*Chen* et al. 1997). Hierbei wurde ein signifikanter Interaktionseffekt zwischen der wahrgenommenen Erwünschtheit positiver und negativer Affektivität und dem durch die Marlowe-Crowne SD-Skala operationalisierten Anerkennungsbedürfnis der Befragten gefunden: Die Angaben der Probanden über ihre eigene Affektivität erwies sich zunehmend stärker in Übereinstimmung mit dem, was sie in der Gesellschaft als anerkannt ansahen, wenn es sich um Befragte mit starkem statt mit schwachem Anerkennungsbedürfnis handelte. Hierbei

wurde ebenfalls die Standardoperationalisierung des TD-Konstruktes herangezogen. Diese Interaktion zwischen den mit der Standardvorgehensweise operationalisierten Erwünschtheitswahrnehmungen und der Stärke des Anerkennungsbedürfnisses konnte dagegen in einer anderen Studie mit Fragen etwa aus dem Bereich der Ehezufriedenheit und über die ethnischen Vorurteile der Befragten nicht festgestellt werden: Beide Faktoren wirken sich in der von *Phillips* und *Clancy* (1972) durchgeführten Untersuchung unabhängig voneinander auf das Antwortverhalten aus. In einer Studie von *Stocké* (2004) wurden die Erwünschtheitswahrnehmungen der Befragten über verschiedene Einstellungen zu Ausländern mit globalen Differenzwerten operationalisiert. Dabei wurde die extrinsisch evaluierte Erwünschtheit der Zustimmung und Ablehnung der insgesamt 10 untersuchten Einstellungsfragen erfasst. Die Ergebnisse haben gezeigt, dass sich die Erwünschtheitswahrnehmungen der Befragten in interviewer-administrierten Befragungen signifikant stärker auf das Antwortverhalten auswirken, verglichen mit der Bedingung einer sichergestellten Privatheit in selbst-administrierten Interviews. Außerdem hat sich dieser Zusammenhang für Befragte mit hohem Anerkennungsbedürfnis als ebenfalls signifikant stärker erwiesen. Die teilweise widersprüchlichen Ergebnisse sind möglicherweise das Ergebnis der unterschiedlich differenzierten Operationalisierung der Erwünschtheitswahrnehmungen in den vorliegenden Studien und der je nach untersuchtem Befragungsthema verschieden starken Verletzung der Neutralitäts- und Monotonieannahme.

2.3 Antwortverzerrungen durch soziale Erwünschtheit bei den untersuchten Befragungsthemen

Die Ergebnisse einiger Studien erbringen Belege für die Existenz, Stärke und Richtung von Antwortverzerrungen durch soziale Erwünschtheit bei den Themen „Alkoholkonsum“ und „Lebensglück“. Dagegen sind uns für das Ausmaß des Fernsehkonsums und der Häufigkeit des Blutspendens keine Untersuchungen über den hier zu erwartenden SD-Bias bekannt. In einer Studie von *Blair* und Kollegen (1977) wurde der auf der Grundlage einer für die Vereinigten Staaten bevölkerungsrepräsentativen Befragung berechnete Alkoholkonsum mit Daten über den landesweiten Alkoholverkauf verglichen. Nach den Ergebnissen der Studie wird der tatsächliche Alkoholkonsum in der Bevölkerung durch Umfragedaten deutlich unterschätzt. Die Autoren ziehen die Schlussfolgerung, dass Befragungsdaten durch Einflüsse sozialer Erwünschtheit in Richtung auf zu geringe Angaben über den Alkoholkonsum verzerrt sind. In einer anderen Studie wurden Soldaten in einer isolierten militärischen Polarstation über das Ausmaß ihres Alkoholkonsums befragt (*Embree* und *Whitehead* 1991). Der auf diesen Angaben beruhende Gesamtkonsum konnte mit den offiziellen Verkaufsdaten der Basis verglichen werden. Je nach Art der Frageformulierung, in der Studie wurden auch der Einfluss unterschiedlicher Referenzperioden und der Einfluss einer Vignettentechnik überprüft, haben die Befragten einen täglichen Alkoholkonsum von .96 bis 2.66 Getränken an-

gegeben. Bei einem tatsächlichen Konsum von 3.36 Getränken täglich, muss von einer Antwortverzerrung in Richtung auf zu geringe Angaben ausgegangen werden.

Bei der Untersuchung der Stärke und Richtung des SD-Bias wird häufig, mangels objektiver Validierungsdaten, auf indirekte Evidenz zurückgegriffen. Hierbei wird geprüft, ob und in welchem Ausmaß sich theoretisch prognostizierte Determinanten des SD-Bias auf die Wahrscheinlichkeit einer Selbstzuschreibung der untersuchten Merkmale auswirken. So wird in einigen Studien die Privatheit der Antwortsituation experimentell variiert und aus den beobachteten Antwortunterschieden Rückschlüsse auf das Ausmaß und die Richtung des SD-Bias gezogen. Da Antwortverzerrungen durch das Bedürfnis nach positiver Selbstdarstellung und Verhaltensbestätigung durch die soziale Umwelt („other deception“) nur bei Wahrnehmbarkeit und daher Sanktionierbarkeit der Angaben erwartet werden, können die Angaben in privaten Interviewsituationen als eine bessere Annäherung an die „wahren“ Werte der Befragten angesehen werden.²

In einer Untersuchung wurden unterschiedliche Aspekte des Alkoholgenusses, wie etwa das Ausmaß des regelmäßigen Konsums oder ob die Befragten jemals ein „Alkoholproblem“ hatten, entweder durch interviewer- oder selbst-administrierte Befragungen erhoben (*Mangione* et al. 1982). Bei den insgesamt 14 untersuchten Aspekten des Alkoholkonsums hat sich nur bei zwei Fragestellungen die Privatheit der Antwortsituation signifikant auf die Randverteilungen der Angaben ausgewirkt. Ähnlich geringe Einflüsse der mit dem Erhebungsmodus verbundenen Privatheit finden sich in einer Studie von *Aquilino* (1994), in der der Alkoholkonsum für verschiedene Referenzperioden und die Häufigkeit der Trunkenheit im vergangenen Jahr durch Verwendung selbst-administrierter Fragebögen, interviewer-administrierter persönlicher Befragungen oder durch Telefoninterviews erfasst wurden. Der Vergleich des Antwortverhaltens zwischen den Modi hat gezeigt, dass in selbst-administrierten Interviews der höchste Alkoholkonsum im vergangenen Monat und am häufigsten eine Trunkenheit von mehr als einmal die Woche angegeben wurden. Insgesamt hat die Antwortprivatheit allerdings in der Gesamtstichprobe nur einen schwachen Effekt, so dass die Autoren nur von einem schwachen SD-Bias in Richtung einer Unterschätzung des Alkoholkonsums ausgehen. Die gleiche Studie findet auch, dass die Angaben von Afroamerikanern, verglichen mit dem Rest der Stichprobe, signifikant stärker durch die Privatheitsbedingungen beeinflusst wurden. In einer anderen Studie mit einer bevölkerungsrepräsentativen Stichprobe wurden ebenfalls die Angaben über das Ausmaß des Alkoholkonsums unter der Bedingung interviewer- und selbst-administrierter Interviews verglichen (*Turner* et al. 1992). Hier fanden sich ebenfalls nur geringe und statistisch nicht signifikante Unterschiede nach der Privatheit der Antwortsituation. In einer weiteren Studie vergleichen *Tur-*

² Für eine Diskussion der Unterscheidung zwischen der extrinsischen, auf die soziale Umwelt gerichteten Motivation zur Fremdtäuschung („other deception“) und der intrinsisch motivierten Tendenz zur positiven Selbsttäuschung („self deception“), vergleiche *Paulhus* und *Reid* 1991).

ner und Kollegen (1998) mit einer bevölkerungsrepräsentativen Stichprobe die Angaben darüber, ob die Befragten Alkohol trinken oder nicht, bei einer selbstadministrierten Audio-CASI Prozedur (computeradministrierte Interviews, bei denen der aufgezeichnete Fragetext über Kopfhörer vorgespielt wird) und einem selbstadministrierten Papierfragebogen. Der anwesende Interviewer kann zwar im Nachhinein den Papierfragebogen einsehen, nicht aber die in den Laptopcomputer eingegebenen Angaben. Somit wird hier nur die Anonymität der Antworten variiert. Es finden sich für den Alkoholkonsum im letzten Jahr keine und für den wöchentlichen Alkoholkonsum nur marginal signifikante Effekte der Experimentalbedingungen. Die Untersuchung von *Locke* und *Gilbert* (1995) hat untersucht, ob sich die Angaben einer Stichprobe von Studierenden über ihren Alkoholkonsum, erfasst durch den „Drinking Habits Questionnaire“, dann unterscheidet, wenn die Daten entweder durch selbst-administrierte Papierfragebögen, selbst-administrierte computergestützte Interviews oder mit einer mündlich-persönlichen Befragung erfasst wurden. Es hat sich gezeigt, dass sich der selbstberichtete Alkoholkonsum nicht nach dem Modus der Befragung unterscheidet. In der Studie von *Waterton* und *Duffy* (1984) wurde im Gegensatz zu den vorher dargestellten Studien in selbst-administrierten Interviews signifikant höhere Angaben über den Konsum von Bier, Wein und Spirituosen beobachtet.

Die Privatheit der Antwortsituation unterscheidet sich auch zwischen Interviews unter Verwendung des sogenannten „Randomized Response“ (RR)-Verfahrens und konventionellen interviewer-administrierten Befragungen. Bei diesem Verfahren kodieren die Befragten ihre Antworten nach einem vorgegebenen Zufallsprozess, so dass sich deren Gesamtverteilung zwar nachträglich ermitteln lässt, die individuellen Antwortinhalte durch die Interviewer jedoch nicht wahrnehmbar sind. In einer Meta-Analyse berichten *Umesh* und *Peterson* (1991) von drei Untersuchungen, in denen die im Rahmen des RR-Verfahrens rekonstruierten Antwortverteilungen über den Alkoholkonsum der Befragten signifikant höher liegen, verglichen mit jenen, die bei einer öffentlichen interviewer-administrierten Befragungsart beobachtet wurden.

Ein anderer, häufig verwendeter Indikator für die Richtung und Stärke des vorliegenden SD-Bias ist die Korrelation zwischen dem mit unterschiedlichen „social desirability“- oder kurz SD-Skalen erfassten Bedürfnis der Befragten nach sozialer Anerkennung und dem Antwortverhalten beim untersuchten Befragungsgegenstand: je stärker sich Befragte mit zunehmendem Anerkennungsbedürfnis das betreffende Merkmal selbst zuschreiben bzw. dessen Existenz leugnen, desto stärkere Antwortverzerrungen werden angenommen. Dabei ist das Vorzeichen der beobachteten Zusammenhänge der Indikator für die Richtung des SD-Bias. In einer Studie von *Cox* und Kollegen (1994) wird das Antwortverhalten der Befragten beim „Michigan Alcoholism Screening Test“ mit den durch die Marlow-Crowne SD-Skala erfassten individuellen Unterschieden im Bedürfnis nach sozialer Anerkennung in Beziehung gesetzt. Die Ergebnisse der Studie zeigen, dass die Befragten mit steigendem Anerkennungsbedürfnis immer geringere Tendenzen zum Alkoholmissbrauch berichten: je

nach Geschlecht der Befragten werden negative Zusammenhänge zwischen $r=-0.07$ und $r=-0.46$ beobachtet. Auch der Zusammenhang zwischen dem Bedürfnis der Befragten nach sozialer Anerkennung und deren Antwortverhalten über das eigene Lebensglück wurde empirisch untersucht (*Kozma* und *Stones* 1988).³ Dabei wurden positive Korrelationen zwischen $r=.21$ und $r=.40$ zwischen den Werten der Befragten auf der Marlow-Crowne SD-Skala und den drei untersuchten Instrumente zur Erfassung des Lebensglücks gefunden. Demnach ist bei diesem Befragungsthema mit SD-Bias in Richtung auf zu hohe Angaben über das Lebensglück zu rechnen. Auch eine Studie von *Phillips* und *Clancy* (1972) findet eine signifikante Korrelation zwischen den Antworten auf die Frage „Alles zusammengenommen, würden Sie sagen, dass Sie im Moment sehr glücklich, ziemlich glücklich oder nicht allzu glücklich sind?“ und den Werten der Befragten auf einer Kurzversion der Marlow-Crowne SD-Skala. Dieses Ergebnis wurde auch im deutschsprachigen Raum repliziert (*Költringer* et al. 1989). Im Gegensatz zu den bisher berichteten Ergebnissen fand eine Studie von *Laicardi* und Kollegen (2001) keine Hinweise auf die Wirksamkeit des SD-Bias. Das mit der Eysenck-Lügenskala operationalisierte Bedürfnis nach sozialer Anerkennung hat sich als unabhängig von den durch die Salamon und Conte „Life Satisfaction Scale“ erfassten Angaben der Befragten über ihre Lebenszufriedenheit erwiesen.

In einigen Studien wurde das sogenannte „Bogus Pipeline“-Verfahren (BP-Verfahren) zur Analyse der Stärke und Richtung des bei verschiedenen Befragungsthemen wirksamen SD-Bias herangezogen. Hierbei wird den Befragten durch eine aufwendige Apparatur die Fähigkeit des Experimentators zur Kontrolle des Wahrheitsgehaltes der Antworten glaubhaft gemacht. Wegen der Abschreckungswirkung der somit für die Probanden subjektiv hohen Entdeckungswahrscheinlichkeit von Falschangaben, wird unter „Bogus Pipeline“-Bedingungen von wahrheitsgemäßerem Antwortverhalten ausgegangen (*Jones* und *Sigall* 1971). So haben etwa *Tourangeau* und Kollegen (1997) das Antwortverhalten der Probanden über ihren Alkoholkonsum unter BP-Bedingungen mit dem bei einer persönlichen Befragung verglichen. Es hat sich gezeigt, dass die Befragten bei fingierter Wahrheitskontrolle signifikant häufiger angeben, mehr als der Durchschnitt zu trinken (21% vs. 4%) und eher einen übermäßigen Alkoholgenuss zugeben (34% vs. 16%). Eine Meta-Analyse über den selbstberichteten Alkoholkonsum unter BP-Bedingungen findet dagegen sehr inkonsistente Ergebnisse (*Aguinis* et al. 1995). Manche der analysierten Studien fanden unter BP-Bedingungen deutlich höhere Angaben über den Alkoholkonsum, während in anderen Untersuchungen weniger Alkoholgenuss als in konventionellen Befragungen berichtet wurde.

In einigen Studien wurden die Erwünschtheitswahrnehmungen bei den Themen „Alkoholkonsum“ und „Lebensglück“ explizit erforscht und teilweise mit der Selbstzuschreibung

3 Bei den drei verwendeten Skalen handelt es sich um die „Philadelphia Geriatric Center Moral Scale“, die „Memorial University of Newfoundland Scale of Happiness“ und den „Life Satisfaction Index-Z“.

dieser Merkmale durch das Antwortverhalten in Beziehung gesetzt. In der Studie von **Bradburn** und Kollegen (1978) haben die Befragten einer bevölkerungsrepräsentativen Stichprobe angegeben, ob sich die Mehrheit der Menschen in der Gesellschaft unbehaglich fühlen würde, Fragen über ihren Alkoholkonsum zu beantworten. Hiervon sind 10 Prozent der Informanten überzeugt, die selbst auch deutlich geringere Angaben über ihren Bier- und Weinkonsum sowie über den Genuss hochprozentiger alkoholischer Getränke gemacht haben. Dies führt die Autoren zur Schlussfolgerung, dass zumindest bei einem Teil der Befragten ein SD-Bias in Richtung auf zu geringe Konsumangaben vorlag.

Im Rahmen der Wortnormenforschung haben Probanden die persönliche Erwünschtheit der Eigenschaften „glücklich“ und „unglücklich“ auf einer Skala von -100 (außerordentlich wenig erwünschte Eigenschaft) bis $+100$ (außerordentlich erwünschte Eigenschaft) bewertet (**Mecklenbräuker** et al. 1994). Dabei zeigte sich, dass die studentischen Informanten eine glückliche Person mit einem durchschnittlichen Skalenwert von $+10.8$ deutlich positiver einschätzen als eine unglückliche Person mit einem Durchschnittswert von -4.2 . In einer anderen Studie wurde die fremdeingeschätzte Erwünschtheit der Merkmale „glücklich“ und „unglücklich“ auf einer Skala von 1 (extrem unerwünscht) bis 9 (extrem erwünscht) erfasst (**Hampson** et al. 1987). Nach den Ergebnissen haben die Befragten eine glückliche Person mit einem mittleren Skalenwert von 8.0 als von anderen Menschen in der Gesellschaft sehr positiv bewertet wahrgenommen. Dagegen wurde bei einer unglücklichen Person im Durchschnitt eine relativ negative gesellschaftliche Bewertung von 2.3 antizipiert. Der Zusammenhang zwischen dem Antwortverhalten der Befragten und der eingeschätzten sozialen Erwünschtheit des Glücklichseins wurde ebenfalls untersucht (**Phillips** und **Clancy** 1972). Es wurde hierbei nur die Bewertung des Merkmals „happy“ auf einer Skala von 1 bis 9 herangezogen. Die Ergebnisse haben gezeigt, dass sich Befragte mit einer hohen Erwünschtheitswahrnehmung des Lebensglücks auch mit hoher Wahrscheinlichkeit selbst als „glücklich“ bezeichnet haben.

Die vorliegenden Studien finden bezüglich des Alkoholkonsums und des Lebensglücks häufig, allerdings keineswegs immer Hinweise auf das Vorliegen von Antwortverzerrungen durch soziale Erwünschtheit. Diese inkonsistenten Befunde könnten einerseits darauf zurückgeführt werden, dass die für die Entstehung des SD-Bias relevanten Faktoren immer nur unvollständig berücksichtigt wurden. Andererseits könnten die Ursache hierfür auch in einer unangemessenen Erfassung der Erwünschtheitswahrnehmungen gesehen werden: in allen berichteten Studien wird ausschließlich die wenig differenzierte Standardoperationalisierung herangezogen.

3 Empirische Untersuchung

3.1 Stichprobe und Repräsentativität

Die Teilnehmer unserer empirischen Studie entstammen einer lokalen, mehrstufigen Zufallsstichprobe der Wohnbevölkerung einer süddeutschen Großstadt. Die Grundgesamtheit der im Spätjahr 1999 durchgeführten Befragung bildeten Bürgerinnen und Bürger mit Lebensmittelpunkt im Stadtgebiet, die zum Befragungszeitpunkt mindestens 18 Jahre alt waren und die deutsche Staatsbürgerschaft besaßen. Durch die Mitarbeiter des Forschungsprojektes wurde mit einer Zufallsbegehung die Haushaltsstichprobe generiert und den Interviewern vorgegeben (Adress-Random). Innerhalb der Haushalte wurde immer die Person für eine Befragung ausgewählt, die vor dem Kontaktzeitpunkt als letzte Geburtstag hatte.

Mit 143 verwertbaren Interviews ließ sich nur eine relativ schlechte Ausschöpfungsquote von 33.3 Prozent realisieren. Ein Vergleich der Zusammensetzung der in der Nettostichprobe repräsentierten Befragten mit den teilweise bekannten und teilweise durch Daten auf einer höheren Gebietsebene angenäherten sozialstrukturellen Merkmale des Auswahlgebietes ergibt die üblicherweise in Umfragedaten vorliegenden Abweichungen.⁴

3.2 Vorgehensweise

Die Befragung erfolgte in allen Fällen als computergestütztes, persönliches Interview bei den Befragten zuhause. Die Interviewer lasen die Fragen laut vor, legten die passende Antwortskala in Papierform vor und gaben dann die Antworten der Probanden in den Laptop-Computer ein. Die Erhebung der Erwünschtheitswahrnehmungen bei den vier untersuchten Merkmalen ist Teil eines durchschnittlich 60-minütigen Interviews, wobei die für die vorliegende Untersuchung relevanten Fragen nach der Erhebung von Erwünschtheitswahrnehmungen über Einstellungen zu Ausländern und dem Demographieteil am Ende des ersten Drittels der Interviews gestellt wurden. Um die Belastung der Befragten zu minimieren, wurde die Gesamtstichprobe mit einem Pseudo-Zufallsgenerator des Befragungsprogramms in zwei gleich große Gruppen aufgeteilt. Jede der Teilgruppen des Populationssplits hat ins-

⁴ So sind in der Stichprobe Personen mit Abitur um 22 und Angestellte um 10 Prozentpunkte überrepräsentiert. Dies führt zu einer Unterrepräsentanz der jeweils anderen Bildungs- und Erwerbsstatusgruppen. In Hinblick auf das Einkommen der Befragten sind in der Stichprobe mit hohem Einkommen überrepräsentiert, wohingegen Personen mit niedrigeren Ausbildungsabschlüssen (Lehre) um 13 und ledige Befragte um 10 Prozentpunkte unterrepräsentiert sind. Dagegen liegen hinsichtlich des Alters, des Geschlechts und der Konfession nur geringe Abweichungen von der Zusammensetzung der Grundgesamtheit vor. Die Vergleichsdaten für die Bildungs- und Ausbildungsabschlüsse, den Familienstand und die repräsentative Einkommensverteilung beziehen sich mangels spezifischer Daten auf die Bevölkerung Westdeutschlands im Dezember 1996 (*Statistisches Bundesamt* 1998). Die Aussagen über die Repräsentativität der im Datensatz vertretenen Geschlechts-, Alters- und Konfessionsstruktur beruhen auf spezifischen Zahlen der Stadt, welche die Grundgesamtheit definiert (*Stadt Mannheim* 1998).

gesamt 15 Fragen über die soziale Erwünschtheit unterschiedlicher Merkmale in einer randomisierten Reihenfolge beantwortet. Die Befragten des ersten Splits wurden neben den hier berichteten Wahrnehmungen über die soziale Erwünschtheit des Alkoholkonsums auch über die antizipierte Bewertung von Vorurteilen und sportlicher Betätigung sowie jener des Führerscheinentzugs und von Vorstrafen befragt (N=71). Die zweite Teilgruppe von Befragten hat neben der im vorliegenden Artikel untersuchten Fragen über die gesellschaftliche Bewertung des Fernsehkonsums, des Blutspendens und des Lebensglücks auch Fragen über die Bewertung einer Parteianhängerschaft sowie von Prüfungsmisserfolgen beantwortet (N=72). Die Auswahl der in unserer Untersuchung hier dargestellten Befragungsthemen erfolgte danach, wie häufig entsprechende Fragen in der Umfrageforschung gestellt werden.

3.3 Operationalisierung der verschiedenen TD-Versionen

Die Befragten haben die verschiedenen Merkmale im Sinne ihrer fremdeingeschätzten sozialen Erwünschtheit beurteilt (vgl. Abschnitt 2.1, oben). Sie wurden aufgefordert ihre Überzeugung darüber zu äußern, in welcher Weise eine Person in der Gesellschaft bewertet wird, die eine schwache, mittlere oder starke Ausprägung auf jeder der vier untersuchten Merkmalsdimensionen aufweist (vgl. die genaue Frageformulierung im Anhang). Die Angaben wurden auf einer bipolaren Antwortskala mit einem Wertebereich zwischen -4 (gesellschaftlich sehr negativ bewertet) bis +4 (gesellschaftlich sehr positiv bewertet) erfasst. Die Befragten haben angegeben, wie eine fiktive Person bewertet wird, die „sehr oft“, „gelegentlich“ oder „nie“ zum Blutspenden geht, ein „sehr“, „teilweise“ oder „überhaupt kein“ glückliches Leben führt, „sehr viel“, „mittelmäßig viel“ oder „nie“ Zeit vor dem Fernseher verbringt und wie eine Person bewertet wird, die „sehr häufig“, gelegentlich“ oder „nie“ Alkohol trinkt.⁵ Die *erste* und unaufwendigste Standardoperationalisierung des TD-Konstruktes besteht in der von den Informanten antizipierten gesellschaftlichen

⁵ Die hier gewählte Erfassung der Erwünschtheitswahrnehmungen bezieht sich auf konkrete Frageformulierungen, die in unterschiedlichen Umfragen in der Vergangenheit verwendet wurden. So wurde etwa im Eurobarometer 1994 (ZA-Studiennummer 2490) die Frage gestellt: „How often have you given blood? Only once/a few times/many times“. Eine ähnliche Frage wurde auch in der Studie „Aktionsgrundlagen der Bundeszentrale für gesundheitliche Aufklärung“ im Jahr 1995 (ZA-Studiennummer 3075) und im National Health Interview Survey 1993 (United States Department of Health and Human Services) gestellt. Im „Quality of Life Survey“ von *Campbell, Converse* und *Rodgers* im Jahr 1971 wurde nach dem Lebensglück der Befragten gefragt: „Asking all things together, how would you say things are these days? Would you say you're very happy, pretty happy or not too happy these days?“ Eine entsprechende Frage zum Lebensglück findet sich auch in der Untersuchung „Promaterielle und postmaterielle Lebensstile“ aus dem Jahr 1991 (ZA-Studiennummer 2421). Auch die Erwünschtheitswahrnehmungen des Fernsehkonsums entsprechen Frageformulierungen in früheren Umfragen. So wurde in der Vorwahlstudie des Jahres 1994 (ZA-Studiennummer 2599) gefragt: „Wie oft schauen Sie Fernsehen? Häufig/Manchmal/Nie“. Die gleiche Frage wurde auch in der Untersuchung „Gleichberechtigung von Frauen und Männern“ 1995 (ZA-Studiennummer 3028) gestellt. Nach dem Alkoholkonsum wurde im „National Health Interview Survey, 1988: Alcohol Supplement“ des United States Department of Health and Human Services ohne Vorgabe von Antwortkategorien gefragt: „On average, how often do you drink any alcoholic beverage?“. Das Eurobarometer 1992 (ZA-Studiennummer 1961) fragte mit sehr detaillierten Antwortkategorien nach dem Alkoholkonsum „How many units do you drink on average per week day

Konstruktes besteht in der von den Informanten antizipierten gesellschaftlichen Bewertung einer starken Ausprägung der unterschiedlichen Merkmalsdimensionen (TD_{STARK}). Als *zweite* Operationalisierung werden auf der Grundlage der Bewertung einer starken und schwachen Ausprägung der verschiedenen Merkmale Differenzwerte gebildet. Zu diesem Zweck werden die Werte beider Rohskalen zuerst auf einen Wertebereich von 0 (gesellschaftlich sehr negativ bewertet) bis +8 (gesellschaftlich sehr positiv bewertet) transformiert und dann die Erwünschtheitsurteile einer schwachen Merkmalsausprägung von denen einer starken Ausprägung abgezogen ($TD_{\text{DIFF}} = TD_{\text{STARK}} - TD_{\text{SCHWACH}}$). Die resultierend globalen Differenzwerte haben einen Wertebereich von -8 (starke Anreize für die Selbstzuschreibung einer schwachen Merkmalsausprägung) und +8 (starke Anreize für die Nennung einer starken Merkmalsausprägung). Bei der dritten und informationsmäßig anspruchsvollsten Umsetzung des TD-Konstruktes werden die Differenzwerte in der wahrgenommenen gesellschaftlichen Bewertung einer starken und mittleren Ausprägung der Merkmalsdimension einerseits und zwischen einer mittleren und schwachen Ausprägung andererseits berechnet ($TD_{\text{ST-MI}} = TD_{\text{STARK}} - TD_{\text{MITTEL}}$; $TD_{\text{MI-SCH}} = TD_{\text{MITTEL}} - TD_{\text{SCHWACH}}$). Jeder der beiden bereichsspezifischen Differenzwerte hat ebenfalls einen möglichen Wertebereich von -8 (starke Anreize zugunsten der Nennung einer mittleren bzw. schwachen Merkmalsausprägung) bis +8 (starke Anreize zur Nennung einer starken bzw. mittleren Merkmalsausprägung). Auch hier wurden die Skalenwerte vor der Differenzbildung auf einen Wertebereich von 0 bis 8 transformiert. Da sich die für das gesamte Merkmalskontinuum prognostizierten Anreize durch soziale Erwünschtheit aus der Summe der beiden Teilanreize ergeben, liegt der insgesamt resultierende Wertebereich dieser Operationalisierung zwischen -8 (starke Anreize für die Nennung einer schwachen Merkmalsausprägung) und +8 (starke Anreize für die Nennung einer starken Merkmalsausprägung). Hinsichtlich der Summe der Beträge der beiden bereichsspezifischen Messungen - hierbei handelt es sich dann um eine Operationalisierung des Verzerrungspotentials durch soziale Erwünschtheit ohne Berücksichtigung der Richtung der Anreize - können Gesamtwerte zwischen 0 (keine Anreize durch SD) und 16 (maximales Verzerrungspotential) realisiert werden. Bei den beiden anderen Operationalisierungen stimmt der Wertebereich des erfassten Verzerrungspotentials dagegen mit dem Betrag der richtungsgebundenen Werte überein (TD_{STARK} : [0/4]; TD_{DIFF} : [0/8]). Dieser Unterschied geht darauf zurück, dass die bereichsspezifischen Differenzwerte die Möglichkeit nicht-linearer Erwünschtheitsverläufe über das Merkmalskontinuum erfassen.

Die drei Operationalisierungen des TD-Konstruktes haben auch, unter Berücksichtigung der Richtung des prognostizierten SD-Bias, formal unterschiedliche Wertebereiche: TD_{STARK} [-4/+4], TD_{DIFF} [-8/+8] und $(TD_{\text{ST-MI}} + TD_{\text{MI-SCH}})$ [-8/+8]. Diese Unterschiede gehen inhaltlich darauf zurück, dass die möglichen Anreize durch soziale Erwünschtheit verschieden voll-

from the time you get up till the time you go to bed?“. Es wurde auch, vergleichbar zu den anderen Themen, danach gefragt, ob die Probanden „häufig“, „gelegentlich“ oder „keinen“ Alkohol konsumieren.

ständig erfasst werden. Ausgehend von der Standardoperationalisierung erweitert sich der Wertebereich bei den globalen Differenzwerten daher von +4 bis -4 auf +8 bis -8, weil hier auch die Möglichkeit einbezogen wird, dass eine schwache Merkmalsausprägung extrem positiv und eine starke Ausprägung extrem negativ beurteilt wird, und umgekehrt. Aus der Logik der Standardoperationalisierung werden dagegen nur Abweichungen vom neutralen Nullpunkt in eine bestimmte Richtung als möglich angesehen. Dies gilt auch für den erweiterten Wertebereich der bereichsspezifischen Differenzwerte. Die bei den Operationalisierungen resultierenden unterschiedlichen Wertebereiche sind somit das Resultat dessen, dass die Neutralitätsannahme und die Monotonieannahme bei den beiden differenzierteren Operationalisierungen aufgegeben wird. In diesen Unterschieden drückt sich somit *nicht* eine mangelnde Vergleichbarkeit der Konstrukte, sondern gerade die erweiterte Fähigkeit der Operationalisierungen zur Erfassung der Erwünschtheitswahrnehmungen aus.

3.4 Ergebnisse

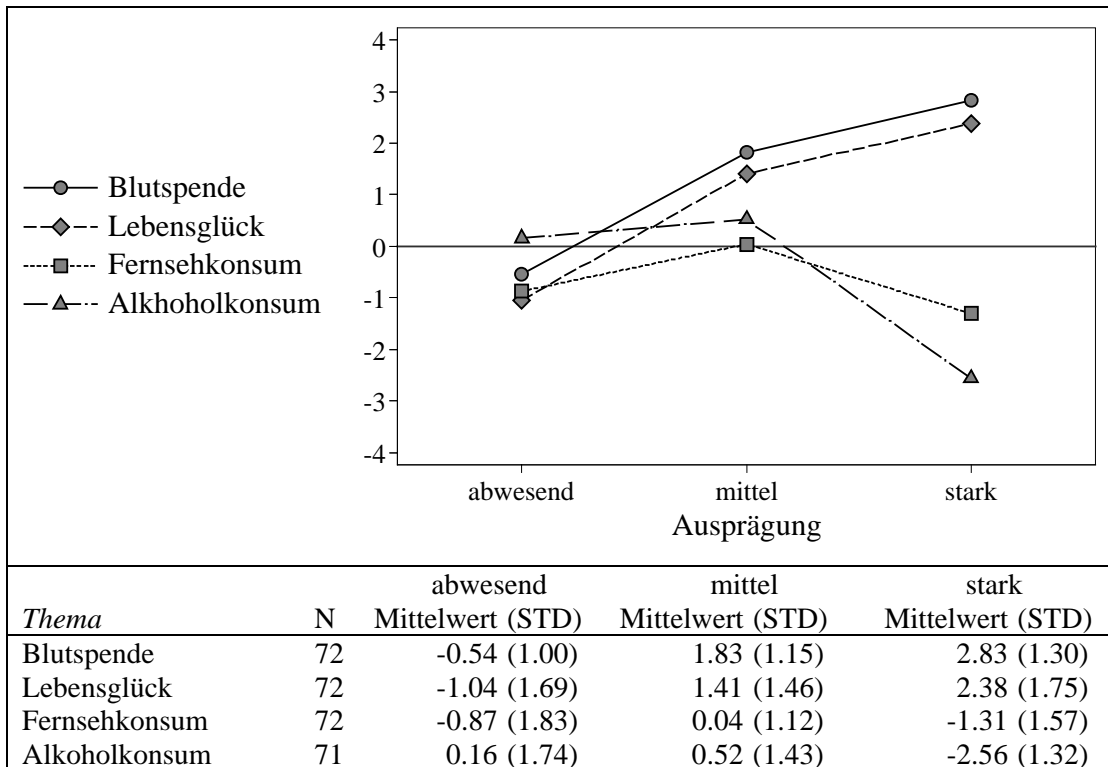
Die Präsentation der Ergebnisse unserer empirischen Analyse erfolgt in vier Schritten. Zuerst werden die durch die drei Operationalisierungen des TD-Konstruktes auf der Aggregatebene der Gesamtstichprobe für die untersuchten Merkmalsdimensionen prognostizierten Anreize durch soziale Erwünschtheit dargestellt. Im zweiten Schritt wird dann die der Standardoperationalisierung zugrunde liegende Annahme einer neutralen Bewertung der schwachen Merkmalsausprägungen empirisch überprüft und die Konsequenzen etwaiger Verstöße gegen diese Neutralitätsannahme analysiert. Im dritten Schritt wird dann getestet, ob sich die bei Verwendung der verschiedenen TD-Operationalisierungen jeweils vorhergesagte Richtung der Anreize durch soziale Erwünschtheit unterscheidet. Dabei wird untersucht, ob die im Rahmen der Standardoperationalisierung und dem Konzept der globalen Differenzwerte gemachte Monotonieannahme der Erwünschtheitsverläufe empirisch erfüllt ist. Im vierten und abschließenden empirischen Teil wird dann gezeigt, ob und bei welchen Befragungsthemen sich für die drei Indikatoren verschiedene Prognosen über die Gesamtstärke des SD-Bias ergeben.

3.4.1 Beschreibung der mit verschiedenen Operationalisierungen des TD-Konzeptes prognostizierten Anreize auf der Aggregatebene

In Abbildung 1 werden die für die unterschiedlichen Ausprägungsstärken der Merkmalsdimensionen beobachteten Rohwerte der sozialen Erwünschtheitswahrnehmungen für die Gesamtstichprobe dargestellt. Dabei wird als erstes Ergebnis deutlich, dass auf der Aggregatebene eine schwache Ausprägung aller untersuchten Merkmale als gesellschaftlich relativ ähnlich und neutral bewertet wahrgenommen wird: Eine Person ohne Lebensglück wird im Durchschnitt mit einem Skalenwert von -1.0, eine Person ohne Fernsehkonsum mit einem Wert von -0.9, eine Person ohne Blutspenden mit einem Wert von -0.5 und eine Person mit

völliger Alkoholabstinenz mit einem Wert von +0.2 bewertet. Die Neutralitätsannahme der Standardoperationalisierung scheint somit weitgehend erfüllt zu sein.

Abbildung 1 Wahrgenommene soziale Erwünschtheit unterschiedlich starker Ausprägungen der Merkmalsdimensionen



Wertebereich der Antwortskalen: -4, „sehr negativ bewertet“ bis +4, „sehr positiv bewertet“.

Es zeigt sich weiterhin, dass ausgehend von der im Aggregat relativ neutralen Bewertung einer fehlenden Merkmalsausprägung ein sehr hohes Lebensglück und insbesondere sehr häufiges Blutspenden als gesellschaftlich deutlich erwünschter wahrgenommen werden. Die durchschnittlich von den Befragten antizipierte Bewertung liegt bei einer starken Ausprägung dieser Merkmale bei +2.4 bzw. +2.8. Dagegen erwarten die Befragten, dass Personen mit intensivem Fernseh- und vor allem hohem Alkoholkonsum eine deutlich negativere Bewertung erfahren. Hier werden durchschnittliche Skalenwerte von -1.3 und -2.6 beobachtet. Während sich die Erwünschtheitswahrnehmungen beim Thema Blutspende und Lebensglück auf der Aggregatebene annähernd linear ansteigend entwickeln, trifft dies für die Verläufe bei den beiden anderen Merkmalsdimensionen nicht zu. Beim Thema „Alkoholkonsum“ lassen sich sowohl für Befragte, die häufig und solche, die nie Alkohol trinken, Anreize für Angaben feststellen, die ein mittleres Konsumniveau anzeigen. Auch beim Fernsehkonsum zeigen die aggregierten Kennzahlen gewisse Anreize zur Angabe einer mittelstarken Ausprägung des Merkmals.

In Tabelle 1 ist die auf Grundlage der drei unterschiedlichen Operationalisierungen des TD-Konstruktes prognostizierte Stärke und Richtung der Anreize zu sozial erwünschtem Antwortverhalten auf der Aggregatebene der Gesamtstichprobe dargestellt. Bei den dargestellten Kennzahlen handelt es sich um Populationsmittelwerte der verschiedenen Anreizskalen bei den unterschiedlichen Befragungsthemen. Ein Vergleich der Standardoperationalisierung (TD_{STARK}) mit der Operationalisierung durch globale Differenzwerte (TD_{DIFF}) zeigt im ersten Schritt, dass beide Indikatoren bei allen Befragungsthemen in der Richtung der prognostizierten Anreize durch soziale Erwünschtheit übereinstimmen: es werden Antwortverzerrungen in Richtung auf ein höheres Lebensglück und häufigere Blutspenden sowie einen geringeren Alkohol- und Fernsehkonsum vorhergesagt. Es lässt sich weiterhin feststellen, dass auf der Grundlage der Differenzwerte signifikant stärkere Anreize zur Angabe häufigerer Blutspenden und einem höheren Lebensglück prognostiziert werden (Blutspende: $t=4.59$, $df=71$; Lebensglück: $t=5.23$, $df=71$; beide: $p < 0.05$). Die weitere Analyse wird zeigen, dass dieser Unterschied auf die Verletzung der bei der Standardoperationalisierung zugrunde liegenden Neutralitätsannahme und der dadurch bei diesen Themen mehrheitlich resultierenden Unterschätzung der tatsächlich vorliegenden Anreize durch soziale Erwünschtheit zurückgeht (vgl. Abschnitt 3.4.2, unten).

Tabelle 1 Durch verschiedene TD-Konstrukte prognostizierte Anreize durch soziale Erwünschtheit auf der Aggregatebene der Merkmale

<i>Thema</i>	TD_{STARK} ϕ (STD)	TD_{DIFF} ϕ (STD)	TD_{ST-MI} ϕ (STD)	TD_{MI-SCH} ϕ (STD)
Blutspende	2.83 (1.30)	3.38 (1.69)	1.01 (1.34)	2.37 (1.38)
Lebensglück	2.38 (1.75)	3.42 (2.78)	0.97 (1.78)	2.45 (2.24)
Fernsehkonsum	-1.31 (1.57)	-0.43 (2.47)	-1.35 (1.52)	0.91 (2.29)
Alkoholkonsum	-2.56 (1.32)	-2.72 (2.40)	-3.08 (1.75)	0.36 (2.63)

Fallzahl: Blutspende, Lebensglück, Fernsehkonsum: $N=72$; Alkoholkonsum: $N=71$.

Beim Merkmal „Alkoholkonsum“ finden sich dagegen keine statistisch abgesicherten Unterschiede in der prognostizierten Anreizstärke zwischen der Standardoperationalisierung und den globalen Differenzwerten ($t=0.76$, $df=70$; $p > 0.1$). Beim Thema „Fernsehkonsum“ werden durch die Standardoperationalisierung sogar signifikant stärkere Effekte sozialer Erwünschtheit vorhergesagt ($t=4.05$, $df=71$; $p < 0.01$). Die aggregierten Kennzahlen der bereichsspezifischen Differenzwerte zeigen, dass bei den beiden letztgenannten Befragungsthemen die Monotonieannahme zumindest bei bestimmten Befragten Gruppen verletzt ist: Die Richtung der aggregierten Anreize dreht sich im niedrigen Ausprägungsbereich der beiden Merkmalsdimensionen um. Da diese nicht-linearen Anreize durch die globalen Differenzwerte nicht berücksichtigt werden und sich (teilweise) gegenseitig aufheben, kann er-

klärt werden, warum die Differenzwerte beim Fernseh- und Alkoholkonsum ein relativ geringes Verzerrungspotential durch den SD-Bias prognostizieren.

Ein weiteres interessantes Ergebnis ist, dass sich nach den bereichsspezifischen Differenzwerten die Anreize durch soziale Erwünschtheit bei den Themen „Blutspende“ und „Lebensglück“ stark zwischen dem Bereich einer starken und schwachen Merkmalsausprägung unterscheiden. Die Befragten nehmen bei beiden Themen mehr als doppelt so starke und damit statistisch signifikant stärkere Anreize zur Nennung einer mittleren statt einer schwachen Merkmalsausprägung wahr, als dies für die Nennung einer starken im Vergleich zu einer mittleren Ausprägung der Fall ist (Blutspende: $t=5.46$, $df=71$; Lebensglück: $t=4.28$, $df=71$; beide: $p < 0.05$). Demnach wird vor allem das Antwortverhalten von Befragten mit sehr geringem Lebensglück und das von denen, die nie zum Blutspenden gehen, durch Anreize zu sozial erwünschtem Antwortverhalten beeinflusst.

3.4.2 Verletzung der Neutralitätsannahme und Konsequenzen für die Validität der Standardoperationalisierung

Die Validität der empirisch unaufwendigen Standardoperationalisierung des TD-Konstruktes hängt unter anderem davon ab, dass die hierbei angenommene neutrale Bewertung einer schwachen Merkmalsausprägung erfüllt ist. Im folgenden wird diese Annahme auf der Ebene der individuellen Erwünschtheitswahrnehmungen der Befragten überprüft und die Konsequenzen eventueller Verletzungen der Neutralitätsannahme analysiert. Zu diesem Zweck werden die Bewertungen einer starken und einer schwachen Ausprägung der unterschiedlichen Merkmalsdimensionen miteinander in Beziehung gesetzt (für die Vorhersagen vgl. Abschnitt 2.1).

Die in Tabelle 2 dargestellten Analyseergebnisse zeigen, dass im Durchschnitt über alle vier Befragungsthemen 62.4 Prozent der Befragten eine schwache Ausprägung der Merkmale gesellschaftlich nicht als neutral bewertet wahrnehmen. Dies ist mit 76.4 Prozent beim Thema „Fernsehkonsument“ am stärksten der Fall, wohingegen die Neutralitätsannahme beim Thema „Blutspenden“ mit einem Anteil von 33.3 Prozent noch am wenigsten verletzt ist. Die beiden anderen Themen weisen bei ungefähr 70 Prozent der Befragten Verletzungen der Neutralitätsannahme auf. Die Abweichung der sozialen Erwünschtheit der schwachen Merkmalsausprägungen von einer neutralen Bewertung führt beim Thema „Blutspende“ in 91.8 und beim „Lebensglück“ in 86.0 Prozent der Fälle zu einer Unterschätzung der vorliegenden Anreize durch soziale Erwünschtheit. Dabei ist die Unterschätzung beim Merkmal „Blutspende“ im Durchschnitt 1.8 und beim Lebensglück 2.2 Punkte der Anreizskala stark.

Tabelle 2 Verletzung der Neutralitätsannahme und deren Konsequenzen für die prognostizierten Anreize durch soziale Erwünschtheit

Thema	Neutralitäts- annahme verletzt		Verletzung führt zu Unterschätzung der Anreizstärke		Verletzung führt zu Überschätzung der Anreizstärke		Verletzung führt zu falschen Schluss- folgerungen über die Richtung der Anreize	
	Anteil ^{a)}	Stärke ^{b)}	Anteil ^{c)}	Stärke ^{b)}	Anteil ^{c)}	Stärke ^{b)}	Anteil ^{c)}	Stärke ^{b)}
Blutspende	33.3	1.71	91.8	1.80	8.2	0.77	0.0	-
Lebensglück	69.4	2.18	86.0	2.16	8.0	2.00	6.0	2.68
Fernsehkonsum	76.4	2.09	47.2	1.88	34.6	1.89	18.2	3.00
Alkoholkonsum	70.4	1.78	48.0	2.26	52.0	1.35	0.0	-
Mittelwert	62.4	1.94	64.1	2.03	28.6	1.50	7.3	1.42

^{a)} Prozentsatz aller Befragter mit einer bei der schwachen Merkmalsausprägung vom neutralen Nullpunkt abweichenden Erwünschtheitswahrnehmung. ^{b)} Durchschnittlicher Betrag der bei einer schwachen Merkmalsausprägung beobachteten Abweichung vom neutralen Nullpunkt der Skala.

^{c)} Anteil der Befragten mit einer Verletzung der Neutralitätsannahme, bei denen die betreffende Art von Messfehler auftritt. Fallzahl: Blutspende, Lebensglück, Fernsehkonsum: N=72; Alkoholkonsum: N=71.

Bei den Befragungsthemen „Alkohol-„ und „Fernsehkonsum“ sind die Folgen der nicht erfüllten Neutralitätsannahme weniger homogen. Beim Thema „Alkoholkonsum“ führt die Verletzung bei 48.0 Prozent zu einer Unterschätzung und bei 52.0 Prozent zu einer Überschätzung der Anreize zum SD-Bias. Allerdings ist die durchschnittliche Unterschätzung mit 2.3 Skaleneinheiten deutlich stärker als die Überschätzung, die im Mittel nur 1.4 Skaleneinheiten beträgt. Beim Befragungsgegenstand „Fernsehkonsum“ ergeben sich bei einer Verletzung der Neutralitätsannahme in 47.2 Prozent der Fälle eine Unterschätzung und bei 34.6 Prozent eine Überschätzung der vorliegenden Anreize durch soziale Erwünschtheit. Bei den letztgenannten Themen heben sich somit die Messfehler in entgegengesetzter Richtung in gewissem Umfang auf der Aggregatebene gegenseitig auf. Die Verletzung der Neutralitätsannahme führt außerdem beim Merkmal „Fernsehkonsum“ bei 18.2 Prozent der Befragten zur Prognose von Anreizen durch soziale Erwünschtheit in die falsche Richtung.

3.4.3 Prognostizierte Richtung der Anreize durch soziale Erwünschtheit auf der Individualebene

Im folgenden Abschnitt wird die Heterogenität bezüglich der in unserer Stichprobe wahrgenommenen Richtung der Anreize durch soziale Erwünschtheit analysiert. Bei der Standardoperationalisierung ist die Richtung durch die Bewertung der starken Ausprägung festgelegt: wird diese positiv bewertet, so liegen Anreize zur Nennung einer starken und bei negativer Bewertung Anreize für die Angabe einer schwachen Ausprägung vor. Für die Operationalisierung mit globalen Differenzwerten ergibt sich die Richtung der Anreize aus dem Vorzeichen dieser Werte. Dies gilt auch für die bereichsspezifische Erfassung, wobei sich hier die Anreizrichtung zwischen den Bereichen unterscheiden kann und dann zu u- oder

umgekehrt u-förmigen Erwünschtheitsverläufen führt. Es wird getestet, ob und gegebenenfalls bei welchen Befragungsthemen die verschiedenen Operationalisierungen der Erwünschtheitswahrnehmungen zu unterschiedlichen Vorhersagen über die Richtung und den Verlauf der Anreize zu Antwortverzerrungen durch soziale Erwünschtheit kommen.

In Tabelle 3 ist die von den unterschiedlichen Operationalisierungen des TD-Konstruktes prognostizierte Richtung der bestehenden Anreize durch soziale Erwünschtheit für die Themen „Blutspende“ und „Lebensglück“ dargestellt. Dabei lässt sich zuerst einmal feststellen, dass alle drei Operationalisierungen des TD-Konstruktes übereinstimmend vorher-sagen, dass die Angaben über die Häufigkeit des Blutspendens potentiell in Richtung auf zu hohe Angaben verzerrt werden: Die Standardoperationalisierung sagt dies für 93.1 Prozent, die Differenzwerte für 94.4 Prozent und die bereichsspezifischen Differenzwerte für 88.9 Prozent der Befragten vorher. Weiterhin prognostizieren alle Operationalisierungen, dass nur um die 5 Prozent der Befragten bei diesem Thema keinerlei Anreize zu Antwortverzerrungen wahrnehmen. Andere Anreiztypen, wie die zugunsten einer schwachen Merkmalsausprägung und umgekehrt u-förmige Anreizprofile, sind wegen der sehr geringen Fallzahlen in diesen Kategorien nicht sinnvoll interpretierbar.

Beim Befragungsthema „Lebensglück“ zeigt sich prinzipiell ein sehr ähnliches Bild wie bei der Blutspende. Hier werden bei Verwendung der Standardoperationalisierung 84.7 Prozent, bei den globalen Differenzwerten 83.3 Prozent und bei den bereichsspezifischen Differenzwerten 76.3 Prozent der Befragten mit einer Anreizwahrnehmung zugunsten der Nennung einer starken Merkmalsausprägung vorhergesagt. Befragte, die keinerlei Anreize für die Nennung einer bestimmten Merkmalsausprägung wahrnehmen, bilden auch hier mit einem Anteil zwischen 5.6 und 9.7 Prozent nur eine kleine Minderheit. Bei der differenzierten, bereichsspezifischen Betrachtung der Erwünschtheitsverläufe wird außerdem eine ebenfalls kleine, 9.7 Prozent starke Gruppe von Befragten mit einer umgekehrt u-förmigen Anreizwahrnehmung vorhergesagt. Diese Befragten gehen davon aus, dass ein mittleres, quasi normales Ausmaß an Lebensglück gesellschaftlich am wünschenswertesten angesehen wird. Da es sich hierbei allerdings absolut gesehen nur um 7 Befragte handelt, verzichten wir bis zu einer Replikation dieses Erwünschtheitsprofils mit einer größeren Stichprobe auf eine substantielle Interpretation dieses Ergebnisses. Insgesamt kann jedenfalls festgestellt werden, dass den verschiedenen Operationalisierungen des TD-Konstruktes bei den Befragungsthemen „Blutspende“ und „Lebensglück“ insgesamt ein hohes Ausmaß an konvergenter Validität hinsichtlich der prognostizierten Richtung der Verzerrungspotentials durch soziale Erwünschtheit zukommt.

Tabelle 3 Durch die TD-Konstrukte prognostizierte Anteile von Befragten mit unterschiedlichen Anreizprofilen bei den Themen „Blutspende“ und „Lebensglück“

<i>Anreize zugunsten einer ...</i>	TD _{STARK}		TD _{DIFF}		TD _{ST-MI} /TD _{MI-SCH}	
	%	N	%	N	%	N
Blutspende						
- starken Merkmalsausprägung	93.1	67	94.4	68	88.9	64
- schwachen Merkmalsausprägung	1.4	1	1.4	1	--	--
- keine Anreize	5.6	4	4.2	3	4.2	3
- u-förmige Anreize	n.v.	n.v.	n.v.	n.v.	--	--
- umgekehrt u-förmige Anreize	n.v.	n.v.	n.v.	n.v.	6.9	5
Gesamt	100.0	72	100.0	72	100.0	72
Lebensglück						
- starken Merkmalsausprägung	84.7	61	83.3	60	76.3	55
- schwachen Merkmalsausprägung	6.9	5	6.9	5	5.6	4
- keine Anreize	8.3	6	9.7	7	5.6	4
- u-förmige Anreize	n.v.	n.v.	n.v.	n.v.	2.8	2
- umgekehrt u-förmige Anreize	n.v.	n.v.	n.v.	n.v.	9.7	7
Gesamt	100.0	72	100.0	72	100.0	72

In Tabelle 4 wird die Richtung der nach den verschiedenen Operationalisierungen des TD-Konstruktes von den Befragten wahrgenommenen Anreize durch soziale Erwünschtheit für die Themen „Fernsehkonsument“ und „Alkoholkonsum“ dargestellt. Hierbei zeigt sich auf der Ebene der individuellen Befragten ein beträchtliches Ausmaß an Heterogenität. So lässt sich feststellen, dass die Standardoperationalisierung des TD-Konstruktes für eine Mehrheit von 72.2 Prozent der Befragten Anreize zugunsten der Angabe eines sehr geringen Ausmaßes an Fernsehkonsums prognostiziert. Für 9.7 Prozent wird eine umgekehrte Prognose gemacht, während 18.1 Prozent keinerlei Anreize durch soziale Erwünschtheit zugeschrieben wird. Bei der Verwendung der globalen Differenzwerte ändert sich die Prognose dergestalt, dass für deutlich mehr Befragte, nämlich 31.9 Prozent, Anreize für die Nennung eines hohen Ausmaßes an Fernsehkonsums vorhergesagt wird. Dieser Unterschied geht auf die durch die Verletzung der Neutralitätsannahme bei diesem Thema im Abschnitt 3.4.2 festgestellten Fehlprognosen der Anreizrichtung zurück. Von besonderer Bedeutung ist jedoch das Ergebnis bei Verwendung der bereichsspezifischen Differenzwerte, wonach 43.1 Prozent der Befragten Anreize zur Nennung eines mittelstarken Ausmaßes an Fernsehkonsums wahrnehmen. In dieser Gruppe wird sowohl ein hoher Fernsehkonsum als auch ein völliger Verzicht als gesellschaftlich weniger anerkannt angesehen. Gleichzeitig nehmen aber auch 31.9 Prozent konsistent Anreize für die Angabe eines vollständigen Fernsehverzichtes wahr, und 18.1 gehen davon aus, dass ein intensiver Fernsehkonsum das höchste Maß an sozialer Anerkennung genießt.

Tabelle 4 Durch die TD-Konstrukte prognostizierte Anteile von Befragten mit unterschiedlichen Anreizprofilen bei den Themen „Fernsehkonsument“ und „Alkoholkonsum“

<i>Anreize zugunsten einer ...</i>	TD _{STARK} %		TD _{DIFF} %		TD _{ST-MI} /TD _{MI-SCH} %	
Fernsehkonsument						
- starken Merkmalsausprägung	9.7	7	31.9	23	18.1	13
- schwachen Merkmalsausprägung	72.2	52	48.6	35	31.9	23
- keine Anreize	18.1	13	19.4	14	4.1	3
- u-förmige Anreize	n.v.	n.v.	n.v.	n.v.	2.8	2
- umgekehrt u-förmige Anreize	n.v.	n.v.	n.v.	n.v.	43.1	31
Gesamt	100.0	72	100.0	72	100.0	72
Alkoholkonsum						
- starken Merkmalsausprägung	1.4	1	1.4	1	1.4	1
- schwachen Merkmalsausprägung	93.0	66	91.5	65	49.3	35
- keine Anreize	5.6	4	7.0	5	1.4	1
- u-förmige Anreize	n.v.	n.v.	n.v.	n.v.	--	--
- umgekehrt u-förmige Anreize	n.v.	n.v.	n.v.	n.v.	47.9	34
Gesamt	100.0	71	100.0	71	100.0	71

Beim Thema „Alkoholkonsum“ prognostiziert die Standardoperationalisierung für 93.0 Prozent und die globalen Differenzwerte für 91.5 Prozent der Befragten Antwortverzerrungen in Richtung auf ein Antwortverhalten, in dem sich ein vollständiger Verzicht auf Alkoholenuss ausdrückt. Bei beiden Anreizindikatoren werden außerdem für 5.6 bzw. 7.0 Prozent der Befragten keinerlei Anreize zu sozial erwünschtem Antwortverhalten vorhergesagt. Auf der Grundlage bereichsspezifischer Differenzwerte wird aber deutlich, dass mit einem Anteil von 47.9 Prozent fast die Hälfte der Befragten Anreize für die Nennung eines mittelstarken und damit moderaten Alkoholkonsum wahrnimmt. Demnach wird, ähnlich wie beim Fernsehkonsum, weder die vollständige Abstinenz noch ein hoher Konsum als die gesellschaftlich anerkannteste Art des Umgangs mit Alkohol angesehen.

3.4.4 Prognostizierte Gesamtstärke der Anreize durch soziale Erwünschtheit

Im folgenden und abschließenden Teil der Analyse wird nun die im Rahmen der drei unterschiedlichen Operationalisierungen des TD-Konstruktes prognostizierte Gesamtstärke der Anreize durch soziale Erwünschtheit verglichen. Hierzu werden die durchschnittlichen Anreizstärken, die als Mittelwerte der Beträge der bei den verschiedenen Operationalisierungen prognostizierten Verzerrungen definiert wurden (siehe Abschnitt 2.1 bzw. 3.3, oben), herangezogen. Die Richtung der Anreize wird somit vernachlässigt.

Tabelle 5 Durch die verschiedenen Konstrukte prognostizierte Gesamtstärke der Anreize durch soziale Erwünschtheit auf der Individualebene

<i>Thema</i>	TD _{STARK} ^{a)}		TD _{DIFF} ^{b)}		TD _{ST-MI} + TD _{MI-SCH} ^{c)}	
	ϕ	(STD)	ϕ	(STD)	ϕ	(STD)
Blutspende	2.89	(1.17)	3.42	(1.60)	3.62	(1.66)
Lebensglück	2.68	(1.22)	3.81	(2.20)	4.20	(2.10)
Fernsehkonsum	1.69	(1.13)	1.90	(1.61)	3.48	(2.14)
Alkoholkonsum	2.65	(1.14)	2.92	(2.16)	5.06	(2.60)
Gesamt	2.48	(0.65)	3.01	(1.04)	4.09	(1.24)

^{a)} Wertebereich: 0, „keine Anreize“ bis +4, „sehr starke Anreize“; ^{b)} Wertebereich: 0, „keine Anreize“ bis +8, „sehr starke Anreize“; ^{c)} Wertebereich: 0, „keine Anreize“ bis +16, „sehr starke Anreize“. Fallzahl: Blutspende, Lebensglück, Fernsehkonsum: N=72; Alkoholkonsum: N=71.

In Tabelle 5 sind diese durchschnittlich vorhergesagten Anreizstärken durch soziale Erwünschtheit dargestellt. Dabei ist erkennbar, dass bei jedem der Befragungsthemen die Standardoperationalisierung das geringste und die bereichsspezifischen Differenzwerte das höchste Verzerrungspotential durch soziale Erwünschtheit vorhersagt. Die globalen Differenzwerte nehmen dabei in jedem Fall eine Mittelstellung ein. Es kann weiterhin festgestellt werden, dass sich bei den Themen „Blutspende“ und „Lebensglück“ der mit den globalen Differenzwerten verbundene Mehraufwand in der Identifikation eines signifikant höheren Verzerrungspotentials durch soziale Erwünschtheit niederschlägt (Blutspende: $t=4.43$, $df=71$; $p < 0.05$; Lebensglück: $t=6.18$, $df=71$; $p < 0.05$). Dieses Phänomen ist eine Folge der Verletzung der Neutralitätsannahme bei beiden Themen und führt dazu, dass die Standardoperationalisierung die Anreize durch soziale Erwünschtheit unterschätzt (vgl. Abschnitt 3.4.2, oben). Dagegen finden sich bei den Merkmalen „Fernsehkonsum“ und „Alkoholkonsum“ kein statistisch abgesicherter Anstieg der identifizierten Stärke der Anreize durch soziale Erwünschtheit, wenn statt der Standardoperationalisierung globale Differenzwerte herangezogen werden (Fernsehkonsum: $t=1.05$, $df=71$; $p > 0.1$; Alkoholkonsum: $t=1.32$, $df=70$; $p > 0.1$). Bei diesen Themen hat sich in Abschnitt 3.4.2 herausgestellt, dass die Verletzung der Neutralitätsannahme teilweise zu einer Über- und teilweise zu einer Unterschätzung der tatsächlich vorliegenden Anreizstärken führen. Diese Messfehler gleichen sich daher in gewissem Umfang gegenseitig aus.

Ausgehend von den mithilfe der globalen Differenzwerte prognostizierten Anreize durch soziale Erwünschtheit, kann mit dem bei den bereichsspezifischen Differenzwerte betriebenen Mehraufwand bei jedem Befragungsthema ein signifikant höheres Ausmaß an Verzerrungspotential durch soziale Erwünschtheit identifiziert werden. So steigt beim Thema „Blutspende“ der Anreizwert von 3.42 um 0.2 Skaleneinheiten auf 3.62 an ($t=2.13$, $df=71$, $p < 0.05$). Beim Thema „Lebensglück“ beträgt dieser Anstieg von 3.81 auf 4.20 der Anreizskala 0.39 Skaleneinheiten ($t=2.88$, $df=71$; $p < 0.05$). Es ist weiterhin erkennbar, dass der bei der

Erfassung von drei Messpunkten notwendige Mehraufwand insbesondere bei den beiden Themen mit einem substantiellen Ausmaß an nicht linearen Erwünschtheitsverläufen zu einer besseren Identifikation des Verzerrungspotentials durch soziale Erwünschtheit führt. Hier steigt das entdeckte Verzerrungspotential beim Befragungsthema Fernsehkonsum von 1.90 auf 3.48 und damit um 1.58 Skalenpunkte an ($t=5.71$, $df=71$; $p < 0.05$). Beim Thema „Blutspende“ beträgt dieser Anstieg, von 2.92 auf 5.06, sogar 2.14 Punkte der Anreizskala ($t=6.36$, $df=70$; $p < 0.05$). Nimmt man die absolute Höhe der von den bereichsspezifischen Differenzwerten, als der differenziertesten und zuverlässigsten Operationalisierung des SD-Bias, prognostizierten Antwortverzerrungen als Referenzpunkt, so kommt bezüglich der erfassten Gesamtanreize keinem der beiden anderen Indikatoren konvergente Validität zu.

Ein Vergleich des mit bereichsspezifischen Differenzwerten identifizierten Verzerrungspotentials durch soziale Erwünschtheit zwischen den untersuchten Befragungsthemen zeigt, dass bei den Themen „Blutspende“ sowie „Fernsehkonsum“ das geringste und ein ähnliches Verzerrungspotential vorliegt: die Anreizstärken von 3.48 (Fernsehkonsum) und 3.62 (Blutspende) unterscheiden sich nicht signifikant ($t=0.46$, $df=71$; $p > 0.1$). Beim Thema „Alkoholkonsum“ liegt mit einem mittleren Skalenwert von 5.06 das stärkste Verzerrungspotential durch soziale Erwünschtheit vor, das signifikant stärker ist, als das beim zweitsensibelsten Merkmal „Lebensglück“ mit einem Wert von 4.20 beobachtet wird ($t=2.3$, $df=70$; $p < 0.05$).

In Tabelle 6 wird überprüft, in welchem Umfang die verschiedenen Indikatoren zumindest darin übereinstimmen, welche Befragten mehr oder weniger Anreize zu Antwortverzerrungen durch soziale Erwünschtheit wahrnehmen. Dieses weitere Kriterium für die konvergente Validität wird durch die Korrelationen zwischen den drei Indikatoren für jedes Befragungsthema überprüft. Dabei zeigt sich, dass bei den Merkmalsdimensionen „Blutspende“ und „Lebensglück“, nun unabhängig von der Anreizhöhe, mittlere bis starke Zusammenhänge beobachtet werden. Dies trifft insbesondere für die Übereinstimmung zwischen den globalen und bereichsspezifischen Differenzwerten zu: hier finden sich Korrelationen von $r=.87$ bzw. $r=.86$. Demnach kann die weniger aufwendige Operationalisierung mit zwei Messpunkten hier als recht gute Annäherung an die optimale bereichsspezifische Erfassung der Anreize angesehen werden. Für die Standardoperationalisierung ist dies mit Korrelationen zwischen .61 und .65 schon deutlich weniger der Fall.

Tabelle 6 Korrelation zwischen der von der verschiedenen TD-Konstrukten prognostizierten Stärke der Gesamtanreize durch soziale Erwünschtheit (Pearsons r)

<i>Thema</i>	Korrelation von TD _{STARK} mit TD _{DIFF}	Korrelation von TD _{STARK} mit TD _{ST-MI} + TD _{MI-SCH}	Korrelation von TD _{DIFF} mit TD _{ST-MI} + TD _{MI-SCH}
	r	r	r
Blutspende	0.78**	0.61**	0.87**
Lebensglück	0.74**	0.65**	0.86**
Fernsehkonsument	0.27*	0.54**	0.24*
Alkoholkonsum	0.61**	0.58**	0.30*

Signifikanz: ** $p < 0.01$, * $p < 0.05$; Fallzahl: Blutspende, Lebensglück, Fernsehkonsument: N=72; Alkoholkonsum: N=71.

Bei den beiden Befragungsthemen „Fernsehkonsument“ und „Alkoholkonsum“ erweisen sich dagegen die globalen Differenzscores als sehr schlechte Annäherung an die Prognosen der bereichsspezifischen Messung: es finden sich nur schwache Zusammenhänge zwischen $r=.24$ und $r=.30$ zwischen den beiden Indikatoren. Dies kann auf die Vernachlässigung der bei diesen Befragungsthemen für substantielle Teile der Befragtenpopulation nicht-linear verlaufenden Anreizwahrnehmungen zurückgeführt werden. Dagegen erreicht die Standardoperationalisierung, mit Korrelationen mit der bereichsspezifischen Messung zwischen $r=.54$ und $r=.58$, zumindest ein mittleres Korrelationsniveau. Dies trifft auch auf den Zusammenhang von $r=.61$ zwischen den globalen Differenzwerten und der Standardoperationalisierung beim Befragungsthema „Alkoholkonsum“ zu. Dagegen zeigt sich beim Fernsehkonsument hier mit einer Korrelation von $r=.27$ ein sehr geringes Ausmaß an konvergenter Validität. Dies geht darauf zurück, dass die Standardoperationalisierung bei diesem Thema wegen der verletzten Neutralitätsannahme für einen Teil der Befragten falsche Prognosen über die Richtung und damit auch über die Stärke der vorliegenden Anreize macht.

4 Zusammenfassung und Diskussion

Die zentrale Fragestellung unserer Untersuchung besteht darin, wie differenziert die von den Befragten antizipierte gesellschaftliche Bewertung kontinuierlich variierender Merkmale erfasst werden muss, um eine zuverlässige Prognose des Verzerrungspotentials durch soziale Erwünschtheit zu ermöglichen. Damit untersucht der Beitrag eine von drei der im Rahmen der RCT als relevant angesehenen Determinanten des SD-Bias. Es handelt sich hierbei daher um einen besonders hervorgehobenen Faktor, weil dieser nicht nur die Stärke sondern auch die Richtung der zu erwartenden Antwortverzerrungen vorhersagt. Die vorliegende Untersuchung hat als *erstes* und wichtigstes Ergebnis gezeigt, dass bei zwei der vier untersuchten Themen für einen substantiellen Teil der Befragten die in der bisherigen Forschung implizit gemachte Annahme monotoner Erwünschtheitsverläufe nicht erfüllt ist. So nehmen

43 Prozent der Befragten beim Thema „Fernsehkonsum“ und rund die Hälfte beim Thema „Alkoholkonsum“ einen umgekehrt u-förmigen Erwünschtheitsverlauf über das Kontinuum unterschiedlich starker Ausprägungen dieser Merkmale wahr. Demnach antizipieren die Befragten bei einer Nennung eines mittelstarken Fernseh- oder Alkoholkonsums positivere Bewertungsreaktionen als dies bei vollständiger Abstinenz oder intensivem Konsum der Fall ist. Der Vergleich der drei von uns berücksichtigten unterschiedlichen Indikatoren für das Verzerrungspotential durch soziale Erwünschtheit hat bei diesen Themen das Ergebnis erbracht, dass ausschließlich die differenzierteste und daher empirisch aufwendigste Erfassung durch bereichsspezifische Differenzwerte als angemessen angesehen werden kann. Weder die in der bisherigen Forschung annähernd ausschließlich verwendete Bewertung nur einer starken Merkmalsausprägung, noch die globale Erwünschtheitsdifferenz einer starken und schwachen Merkmalsausprägung erfassen unter der Bedingung nicht-monotoner Erwünschtheitsverläufe die wahre Stärke und Richtung der Anreize durch soziale Erwünschtheit.

Als *zweites* wichtiges Ergebnis unserer Untersuchung hat sich bei den Befragungsthemen „Blutspende“ und „Lebensglück“ gezeigt, dass alle drei Indikatoren der Anreize zu sozial erwünschtem Antwortverhalten übereinstimmend Anreize zur Angabe einer starken Merkmalsausprägung vorhersagen: häufigeres Blutspenden und ein größeres Lebensglück werden über die komplette Merkmalsdimension als erwünschter wahrgenommen. Damit kommt allen untersuchten Operationalisierungen konvergente Validität hinsichtlich der Richtung des prognostizierten Verzerrungspotentials durch soziale Erwünschtheit zu. Unsere Analyse hat aber auch gezeigt, dass die der Standardoperationalisierung des TD-Konstruktes implizit zugrundeliegende Annahme einer neutralen Bewertung einer schwachen Merkmalsausprägung im Durchschnitt bei allen Befragungsthemen für 62 Prozent der Befragten nicht erfüllt ist. Dies führt bei den Themen „Blutspende“ und „Lebensglück“ zu einer massiven Unterschätzung des möglichen SD-Bias. Obwohl auch bei der Verwendung globaler Differenzwerte die vorliegenden Anreize durch soziale Erwünschtheit quantitativ nicht vollständig erfasst werden, ist dieser Indikator stark mit den bereichsspezifischen Differenzwerten als zuverlässigste Operationalisierung des TD-Konstruktes korreliert.

Zusammenfassend lassen unsere Ergebnisse die Schlussfolgerung zu, dass die differenzierteste Erfassung des möglichen SD-Bias mit bereichsspezifischen Differenzwerten bei allen untersuchten Befragungsthemen der überlegene Indikator ist. Bei Themen mit nicht-monotonen Erwünschtheitsverläufen ist diese aufwendigste Art der Operationalisierung des TD-Konstruktes unabdingbar. Wegen der bei monotonen Erwünschtheitsverläufen hohen konvergenten Validität globaler Differenzwerte kann unter diesen Umständen die Verwendung dieses sparsameren, weil nur auf zwei Messpunkten beruhenden Indikators als gerechtfertigt angesehen werden.

Es muss betont werden, dass es sich bei den in der vorliegenden Studie untersuchten Erwünschtheitswahrnehmungen nur um einen Indikator für die *möglichen* Antwortverzerrungen durch soziale Erwünschtheit handelt. Ob sich diese tatsächlich in einem SD-Bias niederschlagen, kann erst beurteilt werden, wenn deren Erklärungskraft für das Antwortverhalten untersucht wurde. Hierbei möglicherweise beobachtete Zusammenhänge können allerdings auch auf eine tatsächliche Übereinstimmung zwischen den wahren Merkmalen der Befragten mit gesellschaftlichen Normierungen oder darauf zurückgeführt werden, dass die Befragten ihre Eigenschaftsausprägungen auf ihre Erwünschtheitswahrnehmungen projizieren. Eine abschließende Beurteilung der unterschiedlichen Operationalisierungen der Erwünschtheitswahrnehmungen macht daher den Einbezug weiterer Faktoren, wie etwa der Privatheit der Interviewsituation und dem Anerkennungsbedürfnis der Befragten, notwendig. Nach den Hypothesen der RCT sagen Erwünschtheitswahrnehmungen einen SD-Bias in dem Ausmaß vorher, in dem sich deren Erklärungskraft nach dem Anerkennungsbedürfnis der Befragten und der Privatheit der Antwortsituation unterscheidet. Der Nachweis, dass sich die verschiedenen Operationalisierungen der Erwünschtheitswahrnehmungen im Rahmen einer solchen Gesamterklärung des Antwortverhaltens in ihrer Erklärungskraft unterscheiden, steht noch aus.

Ergebnisse einer Untersuchung von *Stocké* (2004) haben gezeigt, dass die mit globalen Differenzwerten gemessene Erwünschtheit unterschiedlicher Einstellungen zu Ausländern auch bei sichergestellter Privatheit der Antwortsituation, allerdings signifikant schwächer als bei der Wahrnehmbarkeit der Antwortinhalte, mit den Angaben der Befragten über ihre Ausländereinstellung korreliert ist. Dies ist ein Beleg dafür, dass Erwünschtheitswahrnehmungen nicht nur die Anreize zu positiver Selbstdarstellung, sondern in gewissem Umfang die „wahren“ Werte der Befragten erfassen. Es sollte daher auch für die in der vorliegenden Untersuchung erfassten Erwünschtheitswahrnehmungen explizit getestet werden, in welchem Umfang diese tatsächlich einen SD-Bias bei den verschiedenen Befragungsthemen bewirken.

In unserer Studie wurde bei der Erfassung der sozialen Erwünschtheit die jeweils zu bewertende Intensität der analysierten Merkmale verbal beschrieben. Dies birgt eine gewisse Unsicherheit darüber, wie die Befragten diese relativ vagen Quantifizierungen, wie etwa „viel“ oder „mittelmäßig viel“ Zeit vor dem Fernseher zu verbringen, interpretieren (vgl. hierzu etwa *Bradburn* und *Miles* 1979). Eine alternative Vorgehensweise hätte sein können, den Informanten objektiv definierte Quantitäten vorzugeben, und etwa zu fragen, wie eine Person bewertet würde, die „4 Stunden“ oder „2 Stunden“ täglich vor dem Fernseher verbringt. Dabei hätte sich allerdings, neben einer gewissen Überforderung der Befragten durch eine zu genaue Charakterisierung der zu bewertenden Quantität, auch das Problem der angemessenen Auswahl der verschiedenen objektiven Quantitäten ergeben. Wir haben es daher vorgezogen, den Befragten selbst die Interpretation der verbalen Quantifizierungen zu überlassen. Es muss aber betont werden, dass unsere Daten streng genommen für solche Fragefor-

mulierungen Gültigkeit haben, die ähnliche Stimuli wie in unserer Studie zur Definition der Antwortmöglichkeiten verwenden. In jedem Fall aber definieren die Quantifizierungen in unserer Untersuchung eine ordinale Reihenfolge der Merkmalsausprägungen und die Ergebnisse sind jedenfalls in diesem Sinne gültig.

Wegen der Verwendung einer lokal definierten Zufallstichprobe und der vergleichsweise geringen Ausschöpfung können unsere Ergebnisse nicht einfach auf die Verhältnisse einer bevölkerungsrepräsentativen Befragung generalisiert werden. Allerdings betrifft dieses Problem der Generalisierbarkeit vor allem die beobachteten Randverteilungen der Erwünschtheitswahrnehmungen und Anreizstärken. Die Schlussfolgerungen in Bezug auf unser Hauptziel, nämlich dem Vergleich der Validität unterschiedlicher Operationalisierungen der wahrgenommenen Anreize zu sozial erwünschtem Antwortverhalten, sind hiervon nicht betroffen.

Literatur

- Aguinis, Herman; Pierce, Charles A.; Quigley, Brian M.* 1995: Enhancing the Validity of Self-Reported Alcohol and Marijuana Consumption Using a Bogus Pipeline Procedure: A Meta Analytic Review. In: *Basic and Applied Social Psychology* 16: 515-527.
- Aquilino, William S.* 1998: Effects of Interview Mode on Measuring Depression in Younger Adults. In: *Journal of Official Statistics* 14: 15-29.
- Aquilino, William S.* 1994: Interview Mode Effects in Surveys of Drug and Alcohol Use. In: *Public Opinion Quarterly* 58: 210-240.
- Blair, Ed; Sudman, Seymour; Bradburn, Norman M.; Stocking, Carol* 1977: How to Ask Questions About Drinking and Sex: Response Effects in Measuring Consumer Behavior. In: *Journal of Marketing Research* 14: 316-321.
- Bochner, Stephen; van Zyl, Theresa* 1984: Desirability Ratings of 110 Personality-Trait Words. In: *The Journal of Social Psychology* 125: 459-465.
- Bradburn, Norman M.; Miles, Carrie* 1979: Vague Quantifiers. In: *Public Opinion Quarterly* 43: 92-101.
- Bradburn, Norman M.; Sudman, Seymour; Blair, Ed; Stocking, Carol* 1978: Question Threat and Response Bias. In: *Public Opinion Quarterly* 42: 221-234.
- Chen, Peter Y.; Dai, Tina; Spector, Paul E.; Jex, Steve M.* 1997: Relation Between Negative Affectivity and Positive Affectivity: Effects of Judged Desirability of Scale Items and Respondents' Social Desirability. In: *Journal of Personality Assessment* 69: 183-198.
- Cox, Brian J.; Swinson, Richard P.; Drenfeld, David M.; Bourdeau, Danielle* 1994: Brief Report. Social Desirability and Self-Reports of Alcohol Abuse in Anxiety Disorder Patients. In: *Behaviour Research and Therapy* 32: 175-178.

- Crott, Helmut W.; Roßbrucker, Karl** 1974: Erwünschtheit von Eigenschaften in Abhängigkeit von Alter, Geschlecht und Schichtzugehörigkeit. In: Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie 6: 241-261.
- Edwards, Allen L.** 1957: The Social Desirability Variable in Personality Assessment and Research. New York: The Dryden Press.
- Embree, Bryan G.; Whitehead, Paul C.** 1991: Validity and Reliability of Self-Reported Drinking Behavior: Dealing with the Problem of Response Bias. In: Journal of Studies on Alcohol 54: 334-344.
- Epstein, Joan F.; Barker, Peggy R.; Kroutil, Larry A.** 2001: Mode Effects in Self-Reported Mental Health Data. In: Public Opinion Quarterly 65: 529-549.
- Esser, Hartmut** 1991: Die Erklärung systematischer Fehler in Interviews: Befragtenverhalten als "rational choice". S. 59-78 in: **Wittenberg, Reinhard** (Hrsg.), Person - Situation - Institution - Kultur. Günter Büschges zum 65. Geburtstag. Berlin: Duncker & Humblot.
- Gove, Walter R.; Geerken, Michael R.** 1977: Response Bias in Surveys of Mental Health: An Empirical Investigation. In: American Journal of Sociology 82: 1289-1317.
- Hampson, Sarah E.; Goldberg, Lewis R.; John, Oliver P.** 1987: Category-Breadth and Social-Desirability Values for 573 Personality Terms. In: European Journal of Personality 1: 241-258.
- Huang, Chiou-yan; Liao, Hsin-ya; Chang, Sue-Hwang** 1998: Social Desirability and the Clinical Self-Report Inventory: Methodological Reconsideration. In: Journal of Clinical Psychology 54: 517-528.
- Jobe, Jared B.; Pratt, William F.; Tourangeau, Roger; Baldwin, Alison K.; Rasinski, Kenneth A.** 1997: Effects of Interview Mode on Sensitive Questions in a Fertility Survey. S. 311-329 in: **Lyberg, Lars; Biemer, Paul; Collins, Martin; De Leeuw, Edith; Dippo, Cathryn; Schwarz, Norbert; Trewin, Dennis** (Eds.), Survey Measurement and Process Quality. New York: John Wiley.
- Jones, Edward E.; Sigall, Harold** 1971: The bogus pipeline: A New Paradigm for Measuring Affect and Attitude. In: Psychological Bulletin 76: 349-364.
- Költringer, Richard; Gehmacher, Ernst; Klucsarits; Willibald** 1989: Response-Set-Messinstrumente: Zur Kontrolle Systematischer Messfehler in der Umfrageforschung. In: Österreichische Zeitschrift für Soziologie 14: 34-45.
- Kozma, Albert; Stones, M. J.** 1988: Social Desirability in Measures of Subjective Well-Being: Age of Comparisons. In: Social Indicators Research 20: 1-14.
- Laicardi, Caterina; Baldassarri, Francesca; Artistico, Daniele** 2001: Unidimensionality of Life Satisfaction and its Relation to Social Desirability: A Confirmatory Study of a Short Form of the Life Satisfaction Scale. In: Psychological Report 88: 253-261.
- Locke, Shellie D.; Gilbert, Brenda O.** 1995: Method of Psychological Assessment, Self-Disclosure, and Experiential Differences: A Study of Computer, Questionnaire, and Interview Assessment Formats. In: Journal of Social Behavior and Personality 10: 255-263.

- Mangione, Thomas W.; Hingson, Ralph; Barrett, Jane** 1982: Collecting Sensitive Data. A Comparison of Three Survey Strategies. In: *Sociological Methods & Research* 10: 337-346.
- Mecklenbräuker, Silvia; Hager, Willi; Möller, Holger** 1994: Erwünschtheit von 908 Eigenschaften bei Frauen und bei Männern aus der Sicht von Männern und von Frauen. S. 310-328 in: **Hager, Willi; Hasselhorn, Marcus H.** Handbuch deutschsprachiger Wortnormen. Göttingen: Hogrefe.
- Meleddu, Mauro; Guicciardi, Marco** 1998: Self-Knowledge and Social Desirability of Personality Traits. In: *European Journal of Personality* 12: 151-168.
- Nicotera, Anne M.** 1996: An Assessment of the Argumentativeness Scale for Social Desirability Bias. In: *Communication Reports* 9: 23-36.
- Paulhus, Delroy L.; Reid, Douglas B.** 1991: Enhancement and Denial in Socially Desirable Responding. In: *Journal of Personality and Social Psychology* 60: 307-317.
- Phillips, Derek L.; Clancy, Kevin J.** 1970: Response Bias in Field Studies of Mental Illness. In: *American Sociological Review* 35: 503-514.
- Phillips, Derek L.; Clancy, Kevin J.** 1972: Some Effects of "Social Desirability" in Survey Studies. In: *American Journal of Sociology* 77: 921-940.
- Presser, Stanley; Stinson, Linda** 1998: Data Collection Mode and Social Desirability Bias in Self-Reported Religious Attendance. In: *American Sociological Review* 63: 137-145.
- Singer, Eleanor** 1981: Reference Groups and Social Evaluations. S. 66-93 in: **Rosenberg, Morris; Turner, Ralph H.** (Eds.), *Sociological Perspectives*. New York: Basic Books.
- Smith, Seyda T.; Smith, Kyle D.** 1993: Social Desirability of Personality Items as a Predictor of Endorsement: A Cross-Cultural Analysis. In: *Journal of Social Psychology* 133: 43-53.
- Smith, Tom W.** 1992: Discrepancies Between Men and Women in Reporting Number of Sexual Partners: A Summary from Four Countries. In: *Social Biology* 39: 203-211.
- Stadt Mannheim** 1998: Elektronischer Datensatz über die Verteilung der Bevölkerungsmerkmale im Stadtgebiet, Stand 31.12.1998. Mannheim: Statistikstelle.
- Statistisches Bundesamt** 1998: Statistisches Jahrbuch für die Bundesrepublik Deutschland. Stand 31.12.1996. Wiesbaden: Online-Version: http://infob1.uni-mannheim.de/IBplus/sessions/STATJB_DEUT98-72eca49b.html.
- Stocké, Volker** 2004: Entstehungsbedingungen von Antwortverzerrungen durch soziale Erwünschtheit. Ein Vergleich der Prognosen der Rational-Choice Theorie und des Modells der Frame-Selektion. In: *Zeitschrift für Soziologie* (im Druck).
- Tourangeau, Roger; Smith, Tom W.; Rasinski, Kenneth A.** 1997: Motivation to Report Sensitive Behaviors on Surveys: Evidence From a Bogus Pipeline Experiment. In: *Journal of Applied Social Psychology* 27: 209-222.
- Turner, Charles F.; Ku, Leighton.; Rogers, Susan M.; Lindberg, Laura D.; Pleck, Joseph H.; Sonenstein,**

-
- Freya L.** 1998: Adolescent Sexual Behavior, Drug Use, and Violence: Increased Reporting with Computer Survey Technology. In: *Science* 280: 867-873.
- Turner, Charles F.; Lessler, Judith T.; Devore, James W.** 1992: Effects of Mode of Administration and Wording on Reporting of Drug Use. S. 177-219 in: **Turner, Charles F.; Lessler, Judith T.; Gfroerer, Joseph C.** (Eds.), *Survey Measurement of Drug Use. Methodological Studies*. Rockville: National Institute on Drug Abuse.
- Umesh, U. N.; Peterson, Robert A.** 1991: A Critical Evaluation of the Randomized Response Method: Applications, Validation, and Research Agenda. In: *Sociological Methods & Research* 20: 104-138.
- Waterton, Jennifer J.; Duffy, John C.** 1984: A Comparison of Computer Interviewing Techniques and Traditional Methods in the Collection of Self-Report Alcohol Consumption Data in a Field Survey. In: *International Statistical Review* 52: 173-182.
- Wyner, Gordon A.** 1980: Response Errors in Self-Reported Number of Arrests. In: *Sociological Methods & Research* 9: 161-177.

Anhang

Instrument zur Erfassung der Erwünschtheitswahrnehmungen der Befragten

Ich nenne Ihnen gleich eine Liste mit Verhaltensweisen oder Meinungen einer Person. Können Sie mir bitte jeweils sagen, ob diese Person deswegen von der Gesellschaft positiv, neutral oder negativ bewertet wird? Es geht also auch hier nicht darum, für wie wünschenswert Sie persönlich diese Eigenschaften halten, sondern was die Mehrheit der Bevölkerung in Deutschland darüber denkt. (Antwortskala wird vorgelegt). Sie können Ihre Meinung auf der vorliegenden Skala zwischen dem Wert minus 4, ‚gesellschaftlich sehr negativ bewertet‘, und plus vier, ‚gesellschaftlich sehr positiv bewertet‘, abstufen. Nennen Sie mir bitte immer den zutreffenden Skalenwert.

Wie wird in unserer Gesellschaft eine Person bewertet, die ...

- | | |
|--------------------|--|
| | ... sehr oft zum Blutspenden geht. |
| (Blutspenden) | ... gelegentlich zum Blutspenden. |
| | ... nie zum Blutspenden geht. |
| | ... ein sehr glückliches Leben führt. |
| (Lebensglück) | ... ein teilweise glückliches Leben führt. |
| | ... überhaupt kein glückliches Leben führt. |
| | ... sehr viel Zeit vor dem Fernseher verbringt.. |
| (Fernsehkonsument) | ... mittelmäßig viel Zeit vor dem Fernseher verbringt. |
| | ... nie Zeit vor dem Fernseher verbringt. |
| | ... sehr häufig Alkohol trinkt. |
| (Alkoholkonsum) | ... gelegentlich Alkohol trinkt. |
| | ... nie Alkohol trinkt. |

SONDERFORSCHUNGSBereich 504 WORKING PAPER SERIES

Nr.	Author	Title
04-18	Alexander Zimmer	Dominance-Solvable Lattice Games
04-17	Volker Stocké Birgit Becker	DETERMINANTEN UND KONSEQUENZEN DER UMFORAGEEINSTELLUNG. Bewertungsdimensionen unterschiedlicher Umfragesponsoren und die Antwortbereitschaft der Befragten
04-16	Volker Stocké Christian Hunkler	Die angemessene Erfassung der Stärke und Richtung von Anreizen durch soziale Erwünschtheit
04-15	Elena Carletti Vittoria Cerasi Sonja Daltung	Multiple-bank lending: diversification and free-riding in monitoring
04-14	Volker Stocké	The Interdependence of Determinants for the Strength and Direction of Social Desirability Bias in Racial Attitude Surveys
04-13	Mei Wang Paul Fischbeck	Evaluating Lotteries, Risks, and Risk-mitigation Programs – A Comparison of China and the United States
04-12	Alexander Ludwig Torsten Sløk	The relationship between stock prices, house prices and consumption in OECD countries
04-11	Jens Wüstemann	Disclosure Regimes and Corporate Governance
04-10	Peter Albrecht Timo Klett	Referenzpunktbezogene risikoadjustierte Performancemaße: Theoretische Grundlagen
04-09	Alexander Klos	The Investment Horizon and Dynamic Asset Allocation - Some Experimental Evidence
04-08	Peter Albrecht Cemil Kantar Yanying Xiao	Mean Reversion-Effekte auf dem deutschen Aktienmarkt: Statistische Analysen der Entwicklung des DAX-KGV
04-07	Geschäftsstelle	Jahresbericht 2003
04-06	Oliver Kirchkamp	Why are Stabilisations delayed - an experiment with an application to all pay auctions